Investigating Morning Commute Route Choice Behavior Using Global Positioning Systems and Multi-day Travel Data

A Thesis

Presented to the Academic Faculty

By

Hainan Li

In Partial Fulfillment

Of the Requirement for the Degree

Doctor of Philosophy in Civil Engineering

Georgia Institute of Technology

October 2004

To my dearest parents

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Randall Guensler, for providing me this precious opportunity of participating in this challenging research project, for valuable guidance and expert advice during the past five years. I am grateful to the other members of my dissertation committee, Professors Michael Meyer, John Leonard, Karen Dixon and James Tsai, whose suggestions and constructive comments guided me through the research process.

My thanks go to Jennifer Ogle, who provided considerable insight into my Ph.D research and encouraged me all the way through; Professors Patrick McCarthy and Laurie Garrow, who provided guidance and suggestions on modeling issues; Lisa Rosenstein, who helped editing the dissertation. My thanks also go to my office mates and fellow group members here at the Drive Lab of the School of Civil and Environmental Engineering, who offered many helpful suggestions and made my journey for Ph. D enjoyable.

Last but not least, I would like to thank my parents and my husband for their love, encouragement and support. Without them, this dissertation would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

One of the major impediments to developing a larger body of knowledge in travel behavior than we currently have is the lack of sufficient data at very detailed levels. The lack of sufficient data is the result of the inherent complexity of gathering and subsequently analyzing observations of the phenomena of interest. This is particularly true for route choice, a topic on which scant link-by-link data appear to be available, especially at multi-day level. In fact, very little empirical work is based on real world observation. This dissertation studied the factors that influence morning commuters' route choice and route switching based on objective real-world observations of travel behavior during multi-day period.

This dissertation tests the current route choice model assumption that travel time or travel distance is the only factor influencing drivers' route choice decision. Investigation of the objective route choice factors confirms that minimizing travel time, although very important, is not the only factor that impacts route choice. Several other factors have been identified that impact commuters' route choice. This dissertation examines the choice between using single or multiple morning commute routes. The results indicate the strong explanatory power of work schedule flexibility and trip-chaining on the choice of single or multiple commute routes compared to the commuters' socio-demographic characteristics and commute route related attributes. This dissertation also presents an extensive effort in analyzing GPS-based travel behavior data and develops a methodology to subtract route choice information and trip-level travel information from the GPS-based vehicle activity data.

Chapter 1

Introduction

Travel is vital to human welfare. Urban residents must travel from place to place to earn a living and to fulfill other needs. Therefore, considerable investments are made in urban transportation infrastructure to meet the increasing travel demand. Over the past few decades, transportation planners have realized that merely increasing the system capacity through new infrastructure construction is not a viable solution to satisfy the increasing travel demand and at the same time maintain healthy economic growth, conserve energy and preserve the natural environment. Hence, the focus of transportation investments has been shifted from the construction of new roads to the application of new innovative technologies to achieve more efficient use of the transportation systems. To be effective, such investments require, among other things, knowledge of future travel demand.

Travel demand forecasting is the process used by urban and regional transportation planners to predict transportation activities and the resulting demand upon the transportation systems. Demand is calculated based on assumptions dealing with land use, the number and character of the trip makers, and the nature of the transportation system. The travel demand forecasting process estimates the volumes on the transportation system, which can be either present-day volumes on an existing network or forecasted volumes on alternative future systems. A central element of the estimation procedure is a model of the traveler's decision about which route to take given the origin, destination, and travel mode of a trip.

Since Wardrop [1952] published the first major paper on network equilibrium, traveler's route choice behavior, along with its effects on network performance, has been one of the most studied subjects in transportation research. Advanced Traveler Information Systems (ATIS) and new technologies are attracting increasing attention towards understanding and modeling the behavior underlying drivers' route choice. Yet even today, accurate predictions of travelers' route choice behavior remain elusive.

The problem of route choice faced by an automobile driver is complex. First, there are a large number of possible alternative routes through road networks between a origin and destination pair, and there are complex patterns of overlap between the various route alternatives [Antonisse et al., 1989]. Second, the ultimate route choice decision is the result of many factors: travelers' socioeconomic characteristics such as age, gender, income, personality, habits and preference; travelers' driving experience, and familiarity with the transportation network; trip characteristics, including trip purpose, time and location, flexibility in arrival time, availability of alternative routes, and traffic conditions. Third, traffic information influences travelers' route choice decisions, both before the trip and en-route.

The decision-making process of route choice is also a dynamic process. A learning process is central to the driver's cognition as the information acquired through experience of earlier travel choices is processed before the next decision is made. Moreover, the characteristics of each known alternative route do not have the same importance in a driver's final decision. Based on the relative importance of each characteristic, travelers formulate a choice set of sufficiently attractive alternatives. From this set, travelers make their choices. The chosen route is the one that best satisfies their needs and is consistent with their personal constraints and preferences. Finally, inertia also plays a role in choice behavior, dictating that certain thresholds be crossed before drivers change their habitual behavior [Polydoropoulou et al., 1994].

Current route choice models are normally based on the assumption that travelers minimize their travel time or travel distance. The underlying route choice models of conventional traffic assignment procedures typically use a single measure of travel impedance such as travel time, travel distance, or some simple formula of generalized travel cost. Conventional travel demand models also assume that travelers have full information about the network, and choose the best route from all those available.

Collecting objective route choice data is a tedious and time-consuming process. Even though these data are important for exploring many detailed relationships that have a direct impact on the understanding of route choice behavior, route choice data are not included in the traditional travel diary data collection methods. In truth, very little empirical work is based on real world observations given the absence of objective evidence of the actual route choice behavior.

Global Positioning Systems (GPS) technology is now increasingly utilized in transportation research as the technology becomes more accurate and less expensive. Advancements in GPS technology make route choice data collection for travel diary studies and other transportation applications a reality. The improvements are both in data quality and data quantity, as well as the addition of new data elements that were once too burdensome or expensive to capture. Recent developments in Geographic Information Systems (GIS) provide useful tools to manage the large amount of spatial related data captured by GPS units and to obtain route choice information from the raw GPS data [Wolf et al. 2000].

Dissertation Objectives

Urban commute journeys have been of significant interest to transportation researchers because the journey to work places great strain on the urban transportation system due to its temporal peak. Therefore, this study focuses on automotive route choice of morning commute journeys. This dissertation attempts to describe how commuters behave in the real world situation.

The overall goal of the dissertation is to investigate drivers' route choice behavior for automobile trips using objective route choice data captured by GPS and GIS technologies. The dissertation examines the differences between actual behavior and current route choice model assumptions.

The objectives of this dissertation can be outlined as follows:

- Develop methodologies to differentiate morning commute activities from the raw GPS-based vehicle activity data.
- Develop methodologies to extract commute journey level information from the raw GPS-based vehicle activity data, including trip starting and ending positions, travel time and distance, intermediate stop durations and locations.
- Develop map-matching methodologies to extract route choice information from the raw GPS-based vehicle activity data.
- Identify different spatial route deviation patterns and study the relationship between travel distance and route deviation.
- Test the hypothesis that travel time and travel distance are the only significant factors that influence drivers' route choice decisions. Identify other objective-level route choice impact factors.
- Test the existence of a primary route for commute journeys and stability of morning commute route choice for the same driver during a certain time period. Establish models of commuters' decision for using single or multiple commute routes. Explore the effects of route attributes, traffic situation, commute characteristics and individual and household socioeconomic characteristics on morning commuters' route choice behavior.
- Provide recommendations on improving traffic assignment methods based on research findings in this dissertation and indicate future research recommendations.

Research Methodologies

The objectives of this dissertation are fulfilled by the following methodological approaches.

Literature Review

The dissertation reviews previous route choice studies, including the following aspects:

- Data collection methodologies,
- Route choice decision making process,
- Route choice impact factors,
- Impact of traffic information,
- Dynamic behavior of route choice,
- Interrelation among choices of route, departure time and trip-chaining,
- Route choice models, and
- Traffic assignment models.

Data Collection and Processing

The commute behavior data used in this dissertation are field observations collected in an ongoing in-vehicle activity data collection effort known as Commute Atlanta. This dissertation develops a series of Perl scripts to capture trip level travel information from the raw second-by-second GPS data set. The GIS road network data used in this dissertation is based on Georgia DOT Digital Line Graphs (DLG) and Road Characteristics (RC) database. This dissertation develops a linear referencing procedure that can join the RC database and the DLG road network. The dissertation also develops map-matching procedures to translate the participant's trips onto the GIS digital road network and hence extract the route traveled. Demographic and socioeconomic data for the participating households and drivers are compiled from the Commute Atlanta household travel survey. An Access database organizes data from the above sources into a convenient format for future analysis.

Statistical Analysis

This dissertation utilizes general descriptive statistical analysis methods, and discrete choice models and utility theory to identify the significant factors that influence drivers' route choice and route switching, and to achieve the other research objectives mentioned in the previous section. This dissertation performed paired sample t-tests, Chi-square tests, and ANOVA tests to compare means and distributions; conditional logit models with multiple observations to analyze the objective route-level impact factors of the morning commute route choice behavior; and binary logit models to study the morning commuters' choice between using single and multiple commute routes.

Potential Research Contributions

The basic premise in the traffic assignment is the assumption of a rational traveler, i.e. one choosing the route which offers the least perceived individual costs [Ortuzar and Willumsen, 2002]. Although a number of factors are thought to influence the choice of route when driving between two points, the generalized cost expression of the route choice models usually only considers two factors: time and money. Monetary cost is often deemed proportional to travel distance. This dissertation carries out statistical tests that assess the current route choice model assumption, which assumes travel time or travel distance is the only factor influencing drivers' route choice decision. This dissertation also carries out statistical tests to identify the additional factors that appear to influence commuters' route choice decision. These findings are useful in generating more realistic general route cost functions and allocating trips to the appropriate road segments in a road network. Hence, the research results help improving the traffic assignment procedures that are one of the major transportation planning modeling steps.

Better understanding of route choice behavior can help researchers design algorithms that generate routes based on more realistic assumptions and have greater usefulness and appeal to ATIS customers. If the travelers choose their routes based on travel time and many other considerations, such as minimizing the difficulty of the driving task or maximizing the facility continuity, but an ATIS suggests routes based on travel time only, the ATIS will not suggest routes that travelers would consider the most attractive. A better understanding of the factors which influence route choice behavior and their relative importance should greatly aid the transportation system management and planning process.

The combination of public policy and individual choice results in trip-making patterns. Transportation planners seek means of inferring travel patterns from readily-observable variables, such as those available from the Census data. This dissertation studies the route choice dynamic of selecting a single commute route or multiple routes using a discrete model that is probabilistic, and based on evidence of drivers' valuations of a number of route and trip characteristics as well as the commuters' socio-demographic characteristics. Accurate behavior models at the discrete individual level are crucial in order to capture the activities, decisions, and spatial motion of travelers through time. The recent developments of planning models are based on the simulation of the daily activities of individual travelers. The research findings are also useful to improve the route choice behavior rules used in the transportation simulation models such as TRANSIMS.

The author explores and illustrates the applicability of innovative data collection methods in transportation research, specifically in collecting data that support route choice analysis. A key feature of this dissertation is that it is based on data describing the actual routes choice decision made by individual drivers. Deriving commute journey-level information from the huge set of raw data presents a challenge. The author develops methodologies to derive trip level information, including trip start and end time, origin and destination locations, travel time and distance, route choice, stop locations, and trip itinerary. The author develops methodologies to associate the road characteristic database with the GIS road network and the GPS data. The dissertation also discusses data accuracy and processing techniques that must be handled to obtain accurate trip level information from the GPS data set. As the GPS technology has been more widely used in transportation research, these methodologies can help future researches to make more efficient use of GPS-based travel data in applications such as GPS survey, GIS road inventory, vehicle navigation and automatic vehicle location, travel time and traffic studies, travel behavior studies and others.

Dissertation Outline

Following this introductory chapter, Chapter 2 summarizes the relevant literature of previous research in the area of route choice. Chapter 3 describes the data collection effort of the dissertation. Chapter 4 presents methodologies to extract route traveled from second-by-second GPS data. Chapter 5 describes methodologies to extract trip level commute information. Chapter 6 presents a summary of general commute trends for the nation and the Atlanta metro area. Chapter 7 provides detailed evaluation of the spatial deviation patterns of the different routes used by the same commuter. Chapter 8 is an empirical analysis of objective route choice impact factors. Chapter 9 presents research findings on morning commuters' choice between a single and multiple routes. Finally, Chapter 10 provides a summary of the research findings and a list of suggestions for further investigation.

Chapter 2

Literature Review

The problem of route choice for a traveler might be stated as follows: Given the other characteristics of the trip to be made (purpose, time, origin, destination, and mode, for instance), attributes of the alternative routes, and traveler's personal characteristics, choose the "best" route through the transportation network in terms of some criterion [Antonisse, 1989].

This chapter first reviews the studies on the theory of the route choice decision-making process. The second section focuses on route choice and the factors that influence it. The third section reviews route choice behavior under the influence of traffic information system. The fourth section reviews dynamic route choice behavior. The fifth section reviews the interrelation of route choice, departure time, and trip-chaining decisions. The sixth section reviews route choice models. The seventh section focuses on the traffic assignment theories. The last section summarizes the literature review.

Route Choice Decision Making Process

Ben-Akiva et al. [1991] developed the dynamic traveler's decision making framework as shown in Figure 2.1.

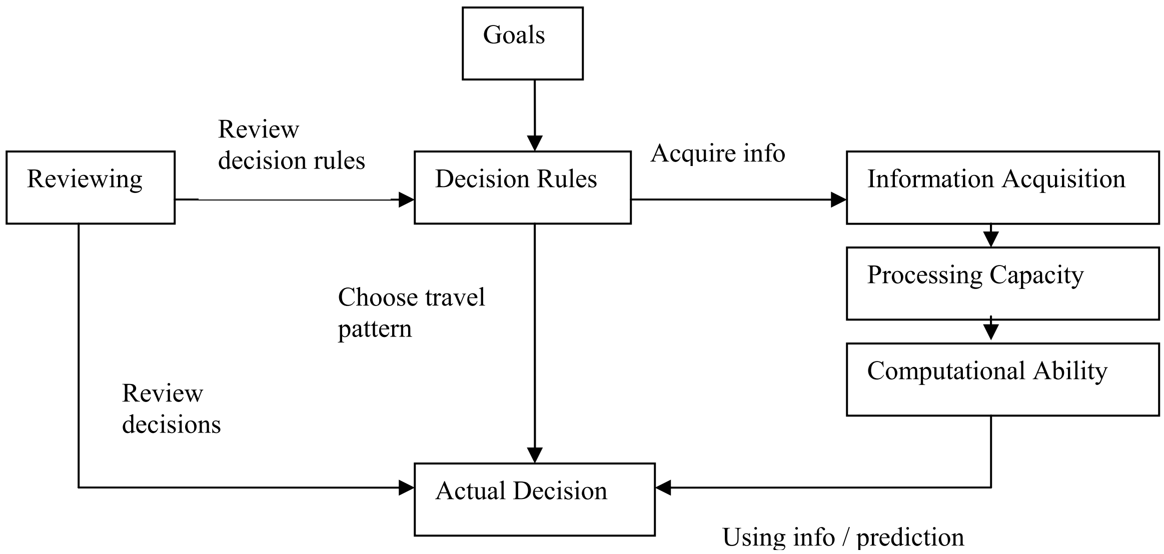


Figure 2.1: Travel Decision Making Process

Source: Ben-Akiva et al. [1991]

Dia [2002] summarized the dynamic travel behavior shown in Figure 2.1 as follows: Drivers seek to travel between an origin and destination during a certain time period at the lowest possible cost in terms of travel distance, travel time or other criteria. Drivers then acquire information about the performance of the road system through direct observation or traffic information. Drivers process and interpret the information in light of their current knowledge and in accordance with their ability to combine and process a variety of information concerning road conditions. The interpretation translates into perceptions of travel time and delay. Perceptions, restrictions and individual characteristics form preferences for certain alternatives (modes, routes, and departure times). The preferences also depend on previously acquired knowledge, stored in the memory, and on certain thresholds whereby motorists only switch from their current path if the improvement in travel time exceeds some threshold level associated with each driver. These preferences result in observable choices that have outcomes (e.g., arrival time at work). If the outcome is satisfactory, the same choice is likely to be repeated on the following trip, forming a commute pattern [Ben-Akiva et al., 1991]. The repetition of a choice in the commute contest also depends on the future or anticipated outcomes. These outcomes also provide feedback to the memory in the form of knowledge updates. In unanticipated delay situations, the anticipated outcomes are often unsatisfactory, triggering review of preferences and changes in normal travel pattern on a real-time and day-to-day basis [Ben-Akiva et al., 1991].

Stern et al. [1993] described the travel behavior from the aspect of man-environment interface paradigm and used the following series: (a) an objective situation, (b) a subjective situation, (c) personal perceptions, and (d) personal decisions. The basic structure of the men-environment interface paradigm is shown in Figure 2.2. The information-wise expanded structure is shown in Figure 2.3.

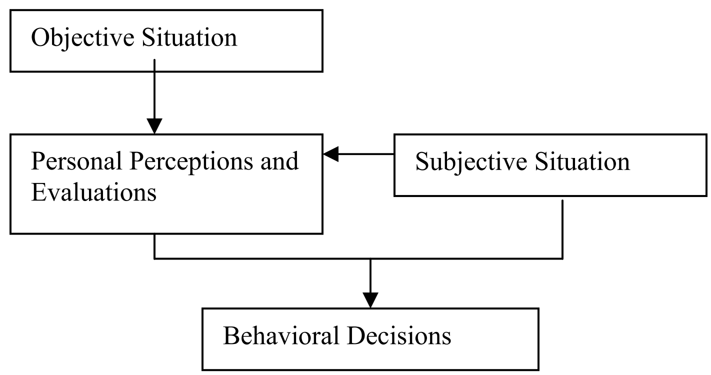


Figure 2.2: Basic Structure of Men-Environment Interface Paradigm of Travel Behavior

Source: Stern et al. [1993]

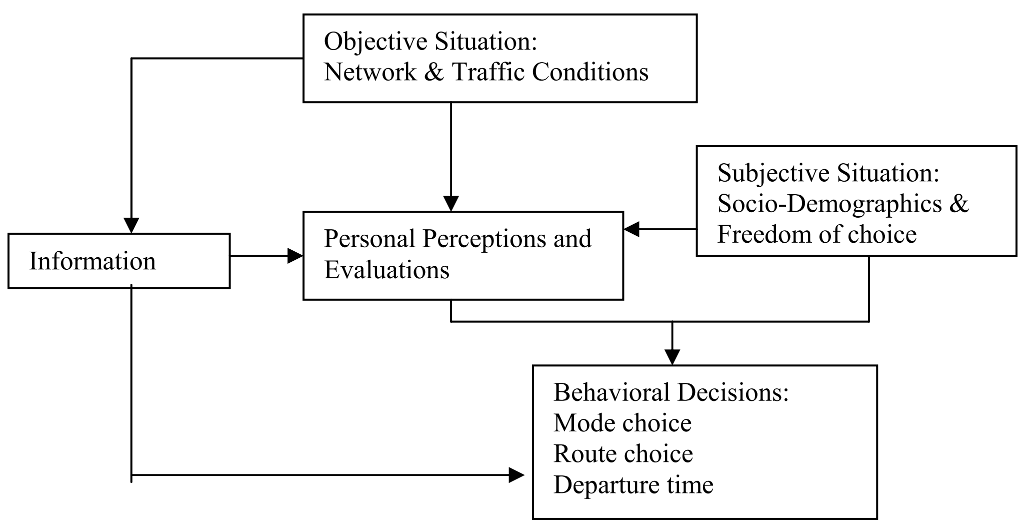


Figure 2.3: Expanded Structure of Men-Environment Interface Paradigm of Travel Behavior

Source: Stern et al. [1993]

Polydoropoulou et al. [1994] mentioned the decision-making process of route choice is a dynamic process. A learning process is central to the driver's cognition as the information acquired through the experience of earlier travel choices is processed before the next decision is made. Moreover, the characteristics of each known alternative route do not have the same importance in a driver's final decision. On the basis of relative importance of impact factors, travelers formulate a choice set of sufficiently attractive alternatives. From this set, travelers make their choices, with the chosen route being the one that best satisfies their needs and are consistent with their personal constraints and preferences. Finally, inertia also plays a role in travel choice behavior, dictating that certain thresholds be crossed before drivers change their habitual behavior.

Antonisse et al. [1989] mentioned a two-step model of route choice decision making process in which step one narrows down the large number of possible route alternatives to a choice set of a few alternatives and step two chooses a route from this choice set based on the characteristics of the trip, drivers, and the attributes of the available alternatives. The route choice process is not a direct and simple derivative of observable characteristics of the transportation network and of the traveler. In step one of the two-step model, limited knowledge of all the opportunities available, and the constraints of the traveler that eliminate some alternatives may generate different choice set for different drivers. In step two, subjective values derived from the objective attributes may be distorted and result in different selection.

A most comprehensive diagram of the decision making process is provided by Bovy and Stern [1990], as shown in Figure 2.4.

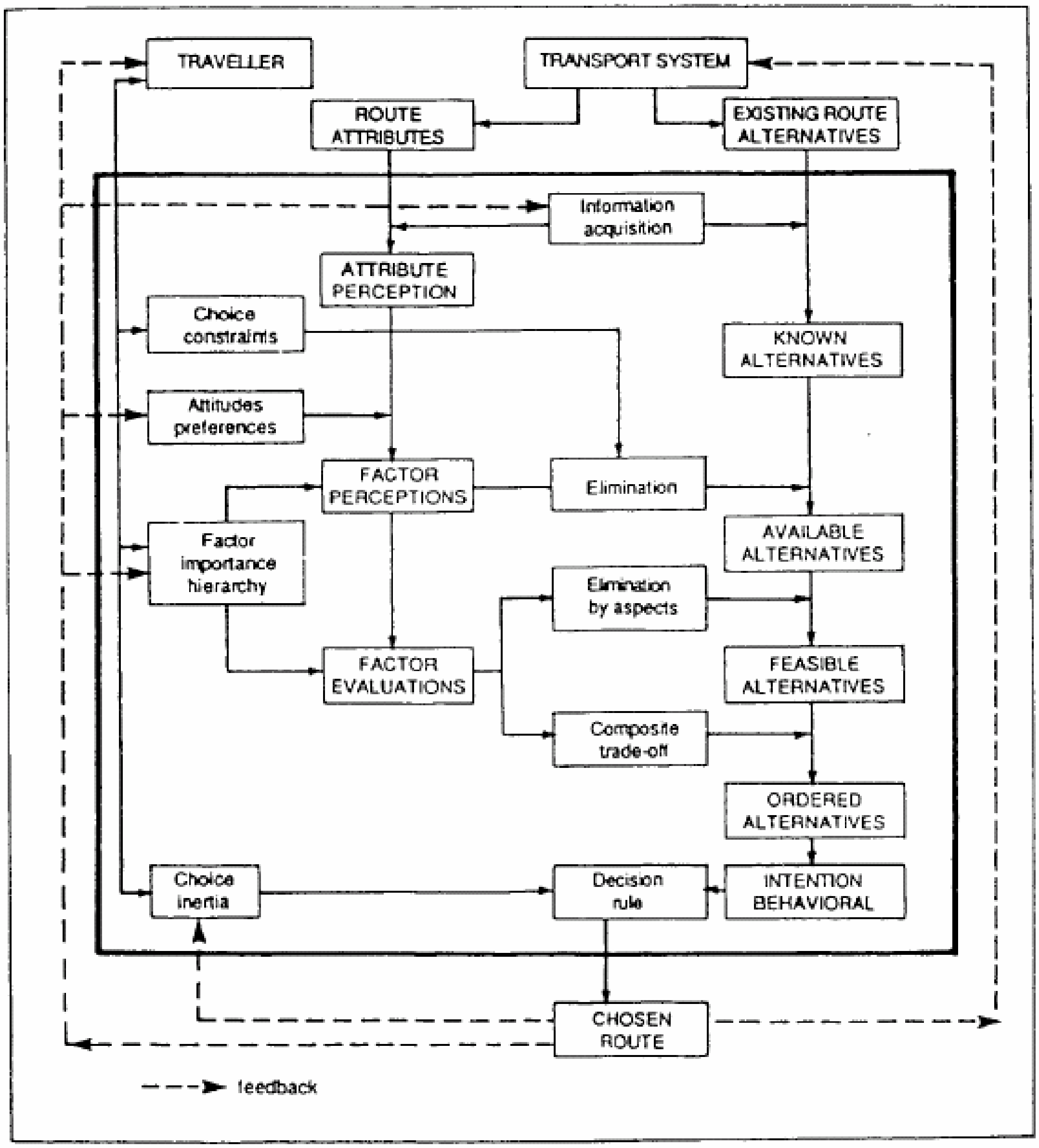


Figure 2.4: Elements of Individual Choice Behavior

Source: Bovy and Stern [1990]

Factors Influence Route Choice

Although the shortest-path routing has been adopted over the years because of its simplicity and linkage with algorithms for generating equilibrium in static traffic assignment models, in real life, driver's routes are likely to significantly deviate from the fastest path [Abdel-Aty et al., 1996]. Empirical research on the route choice behavior shows that drivers use numerous criteria in formulating a route: travel time, number of intersections, traffic safety, traffic lights and other factors. Drivers' experiences, habits, cognitive limits and other behavioral considerations may also produce variations in route selection. Assuming travel time as the sole criterion of route choice may be an unrealistic abstraction of individual driver behavior, and may result in an inaccurate representation of traffic when aggregated at the network level.

Antonisse et al. [1989] summarized previous research findings on specific attributes of routes to which drivers are attracted, which includes travel time, distance, the number of traffic signals, scenery, time or distance on limited-access highways, safety, commercial development, congestion, road quality, and road signing. Stern et al. [1993] summarized all possible factors into four groups of factors, at both the objective and the subjective levels:

- The physical environment, including the network infrastructure which determines, for example, travel possibilities and their characteristics.
- The socio-demographic environment, including the household characteristics like modes of transport chosen, age, and the like. These attributes will affect

cognition and perception of travel opportunities as may also impose constraints on travel.

- The normative environment, including the set of norms, values and concepts derived from society and, in particular, from the immediate surrounding of the traveler.
- The personal environment, comprising of the personality of the decision-maker, which may cause the three factors mentioned above, together forming the objective situation, to be observed subjectively, and the information derived there from to be converted into a decision.

Jan et al. [2000] grouped all the possible factors that influence drivers' route choice behavior into four groups, as shown in Table 2.1.

Table 2.1: Route Choice Factors

Source: Jan et al. [2000]

Traveler	Age, gender, life cycle, income level, education, household structure, race, profession, length of residence, number of drivers in family, number of cars in family, etc.	
Route	Road	Travel time, travel cost, speed limits, waiting time. Type of road, width, length, number of lanes, angularity, intersections, bridges, slopes, etc.
	Traffic	Traffic density, congestion, number of turns, stop signs, and traffic lights, travel speed, probability of accident, reliability and variability in travel time, etc.
	Environment	Aesthetics, land use along route, scenery, easy pick-up/drop-off, safety, parking, etc.
Trip	Trip purpose, time budget, time of the trip, mode use, number of travelers	
Circumstances	Weather conditions, day/night, accident en route, route and traffic information, etc.	

Jackson and Jucker [1981] investigated the impact of a specific measure of reliability, the variability of travel time, on the route-to-work choice through the use of a survey instrument posing hypothetical commute alternatives. The authors suggested that including travel time variability in the impedance function along with travel time might improve the traffic assignment process for two reasons. First, the reliability of transportation systems is considered of prime importance to the traveler. Second, a number of criteria not included in traditional impedance functions, such as the number of stop lights on a route and the safety of that route, may be positively correlated with the variability of travel time measure. The primary conclusion of this research is that a specific measure of reliability, variability of travel time, has an important impact on the route-to-work choice and that this impact varies substantially across individuals, ranging from those who will choose routes that are significantly longer to avoid the possibility of delay to those who are essentially expected value decision makers with regard to commute alternatives. In this study, travel-time is defined as the median of the sample, and travel time reliability is defined as difference between the 90^{th} percentile and the median.

Researchers have also found time-related variables are not the only ones to be considered in traffic assignment procedures. Driving efforts measured with a psychological scale have been found of considerable influence in the individual’s route choice process [Stern et al. 1983]. In their study, driving efforts are measured by the number of turns along an alternative route, and the effect of the number of turns is much stronger for short-time routes.

Route Choice under Travel Information

The provision of real-time travel information is increasingly being recognized as a potential strategy for influencing driver behavior on route choice, trip making, time-of-travel, and mode choice. These systems provide drivers with real-time information about traffic conditions, accident delays, road work, and route guidance from origin to destination. Some of the methods used for providing drivers with this information include traffic information broadcasting, pre-trip electronic route planning, on-board navigation systems, electronic guidance systems, and strategically located variable message signs. The principal aim of these systems is to influence drivers' behavior on route choice and departure time decisions in order to improve mobility and reduce traffic congestion [Dia, 2002].

Route choice is a complex process which becomes even more complicated when traffic information is available to drivers. Understanding this process depends not only on all the usual factors that affect route choice decisions, such as travel time, travel distance, personal preferences, etc., but also on attributes related to the information, such as type, context, spatial and temporal relevance, perceived reliability, etc. Furthermore, existence of on-line traffic information forces decisions to be made in real-time, often under time pressure, and while the driver is primarily occupied with the driving task [Lotan, 1996].

In recent years, an abundance of research has focused on commuters' route choice with an emphasis on how real-time traffic information might affect drivers' route choice behavior and identify distinction between the route choice behavior of drivers who use information and those who do not. A large number of research efforts also have been devoted to the development of route choice model that can include the effect of traveler information.

Polydoropoulou et al. [1994] summarized the influence of traffic information on travelers' pre-trip and en-route behavior, as shown in Figure 2.5.

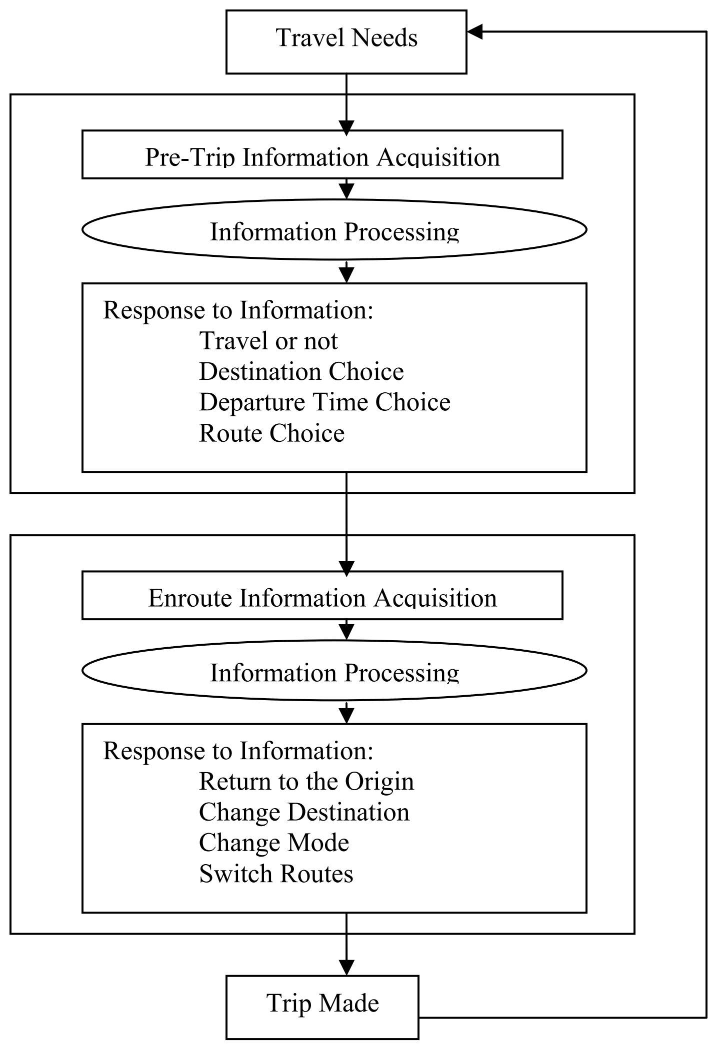


Figure 2.5: Impact of Traffic Information on Travelers' Decision

Source: Polydoropoulou et al. [1994]

Srinivasan and Mahmassani [2000] proposed that the observed route choices in response to information are a consequence of two principal mechanisms operating in the decision-making process: compliance and inertia. These mechanisms reflect the propensities of a user to comply with the ATIS information (best path) and to retain the current path, respectively. The research used route choice data obtained from dynamic interactive simulator experiments. The empirical results strongly support the simultaneous presence of both the compliance and inertia mechanisms in route choice behavior. The results also illustrate that information quality, network loading and day-to-day evolution, level-of-service measures, and trip-makers' prior experience are significant determinants of route choice.

Based on a 1992 computer-aided telephone interview survey of Los Angeles area morning commuters about their route choice behavior and the influence of traffic information, Abdel-Aty et al. [1994] found about 36.5 percent of the respondents listened to traffic reports before leaving their homes, and 51.25 percent listened while driving. Close to 27.6 percent of the respondents listened to traffic reports both at home and en route, and 60.1 percent listen to reports either at home or en route, whereas 39.9 percent never listened to reports. Most respondents who received traffic information perceived the traffic reports to be either very accurate or somewhat accurate.

Stern et al. [1993] carried out two surveys in Sweden and Israel. The survey results show that, on average, two thirds of commuters will change their travel behavior upon receiving traffic information. As observed, about 43 percent of total drivers would change their habitual route upon receiving information about anticipated traffic problems. Only less than 5 percent would change their mode, and about 18 percent would change their departure time. This study identified the discriminating factors between commuters who change and commuters who do not change their travel behavior due to information and found variables which contribute most to this separation. The group means of the discriminate variables provide more information about the profiles of the two separated groups that change behavior and that not. Evidently, commuters who change their travel behavior are more exposed to information, are more sensitive to congestion, and have more freedom of choice. They set up a general model of route change probability in congested situations, as shown in Figure 2.6.

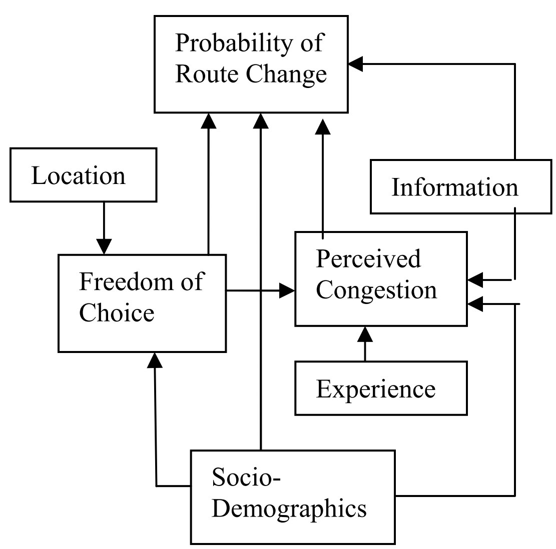


Figure 2.6: Route Change Probability in Congested Situations

Source: Stern et al. [1993]

Khattak et al. [1995] surveyed downtown Chicago automobile commuters during the morning peak period about the effect of traffic information on their route and departure time choice. More than 60 percent of the respondents had used traffic information to modify their travel decisions. Multivariate analysis using the ordered probit model indicated that individuals were more likely to use traffic reports for their route changes if they perceived traffic reports to be accurate and timely, and if they frequently listened to traffic reports. Respondents were more likely to change their departure times if they perceived traffic reports to be accurate and relevant, and if they frequently listened to traffic reports. Providing information of low quality can increase the travel cost more than not providing information at all.

Polydoropoulou et al. [1994] analyzed commuters' route choice behavior in the presence of traffic information based on revealed preference data. Trip characteristics and travelers' perceptions of the relevance and reliability of radio traffic reports were found to be important factors affecting radio traffic information acquisition and its influence on drivers' decisions. The key finding was that en-route diversion is primarily influenced by attitudinal factors and by information acquisition. Moreover, drivers' own observations are important factors affecting route switching.

Emmerink et al. [1995] studied the impact of both radio traffic information and variable message sign information on route choice behavior. Descriptive statistics of these data revealed that over 70 percent of the motorway users are sometimes influenced by these types of information. The analysis show that, 1) women are more reluctant to be influenced by information during the trip than men; 2) commuters tend to be less influenced by information than travelers with other trip purposes; 3) radio traffic information and ATIS information are used in a similar way by travelers.

Based on an intensive literature review by Khattak et al. [1991], the factors found to induce diversion were traffic information, longer travel time on the preferred route, and familiarity with the alternate route. Factors found to inhibit diversion were longer travel times on the alternate route, expected congestion and delay on the alternate route, and traffic stops on the alternate route. Young, male, and unmarried drivers had a higher inclination to switch routes. They also found drivers may express willingness to divert in hypothetical situations, but their actual diversions may be considerably less, perhaps influenced by a variety of situational variables. Median value of delay for diversion and percentage of drivers who divert vary from research to research. Traffic information also has been found to influence diversion. Drivers were more likely to divert to familiar routes, suggesting that cognitive maps of drivers may influence their diversion behavior.

Dynamic Aspect of Route Choice

Mannering [1989] used a Poisson regression to predict the frequency of commuters' route changes per month. He found out that both highway network (e.g., the availability of alternative routes, travel time on the primary route, the level of traffic congestion) and commuters' socioeconomic characteristics play important roles in the frequency of route changes. As a commuter's age increases fewer route changes are made. Unmarried people were found to be more likely to change routes than their married counterparts.

This may be reflecting more risk-seeking or impatient behavior among single commuters, or simply capturing the fact that married commuters may be constrained by the need to take a spouse to work or by some other family responsibilities. Male commuters were found to be more likely than females to change routes.

Mannering and Kim [1994] collected survey data of interstate 5 (I-5) commuters in the Seattle metropolitan area and used ordered logit model to predict the frequency of changing home-to-work routes. Examining specific coefficient estimates, they found that the longer the daily commute time, the higher the frequency of route changes. Commuters indicating that they had considerable flexibility in departure times at home and at work were found to be more frequent route changers. It was also found that the greater the commuters' familiarity with alternative routes, the higher the frequency of route changes. Turning to socioeconomic variables, they found men are more likely to change routes than women, and individuals with low incomes were found to be less likely to change routes frequently.

Mahmassani et al. [1990] conducted a commute survey in Austin, Texas. Their binary logit models relate route switching propensity to four types of factors: geographic and network condition variables, workplace characteristics, individual attributes, and use of information (radio traffic reports). They found out that variables describing the characteristics of the commute itself had a dominant effect relative to workplace rules or individual characteristics. The use of information in the form of radio traffic reports also exerted a strong effect, indicating that regular listeners to traffic reports had a greater propensity to switch routes. The only socio-demographic attribute included in the model was age.

In the survey carried out by Abdel-Aty et al. [1994], only 15.5 percent of the respondents said they used more than one route to work. About 50 percent of the respondents had at least one freeway segment in their primary routes. Secondary routes tend to have more surface streets than primary routes, possibly as alternatives used to avoid congestion on freeways. The most frequent reason for changing routes, cited by 34 percent of respondents, is the traffic that the respondents see on the roads. Individuals with higher incomes tend to report using more than one route to work. Highly educated people tend to use alternate routes.

Interrelation of Route Choice, Departure Time Choice, and Trip-Chaining

Departure time choice and route choice constitute the primary choices available to commuters on a daily basis in response to congestion, incidents, or other situations. In contrast, the time-frames for decisions of mode shifts, telecommuting, residence relocation, and change of work place are comparatively longer.

The work place anchors some of the non-work travel, either in intermediate stops commuters make between home and work or in trips around the workplace. As the empirical evidence pointed out, a secondary role of the commute journey is to provide an opportunity to link non-work travel with the commute itself [Nishii et al., 1989]. Commuting trips are becoming more and more complex as workers incorporate personal, household, and child-care activities into their commutes [Bianco and Lawson, 1996]. Research results from [Mahmassani et al.,1996; Nishii et al.,1989; Davidson,1991; Yalamanchili et al.,1999; and Hanson,1980] highlighted the impact of trip-chaining along the commute on the variability of departure time and route choice decisions. Orski [1989] found that more than 60 percent of the office workers who drive their personal car to work made intermediate stops on the way to or from work at least three times a week. Davidson [1991] also found that employees were twice as likely to make stops on their way home from work as on their way to work from home. Predominant morning chaining was to get gas (45.2%), to go to the bank (22.7%), to go to the dry cleaners (19.4%) to eat (16.4%), and to daycare and school (20%), based on Davidson's study of 42 employer sites and 1845 employees in an employee travel needs survey [Davidson, 1991]. The need to make stops on the way was one frequent reason for changing routes and was cited by 15.5 percent of respondents in [Abdel-Aty et al., 1994].

Lam [2000] developed a theoretical model to analyze commuters' joint decisions of route and departure time in a simple network with two parallel routes. The results confirm the findings of previous studies that commuters shift departure times earlier in response to increase in travel time uncertainty. Pre-trip information may allow later departure times even in cases of heavy traffic condition because it reduces the uncertainty in travel time.

Mahmassani and Stephan [1988] studied the interaction between departure time and route switching in response to experienced congestion in a traffic commuting system using an experimental approach in which actual commuters interact over a period of seven weeks in a simulated commuting system. The study pointed out that the precedence of departure time shifts over route shifting in dealing with experienced unpredicted congestion in the system.

Route Choice Models

Travel behavior in general and route choice behavior in particular can be considered as choosing among discrete, mutually exclusive alternatives [Antonisse et al., 1989]. Discrete choice analysis attaches expressions of attractiveness or utility to each of the available choices options. The utility expression of each alternative generally incorporates information on the attributes that may either add to or detract from its attractiveness. It is then assumed that the decision maker will choose the alternative that is most attractive.

The utility is usually defined as a linear combination of variables, where each variable represents an attribute of the option or the traveler. The relative influence of each attribute, in terms of contribution to the overall satisfaction produced by the alternative, is given by its coefficient. Variables within a utility function can be either generic or alternative-specific in nature. A generic variable is one that is included in every alternative's utility function with exactly the same weight. An alternatives-specific variable, on the other hand, has different weights for different alternatives. Usually one alternative has a priori specified weight of zero to facilitate model estimation.

Conventional microeconomics makes strong assumptions concerning the decision maker's ability to use perfectly all the information available and relevant to the decision and to make a completely rational, consistent decision given this information. A major relaxation of some of these assumptions is possibly through the introduction of the concept of random utility. These models recognize that, in practice, people do not always choose the "objectively best" course of action, nor do they necessarily exhibit consistent choices over time. That is, random utility theory still assumes that an individual will choose that alternative which appears to maximize his or her utility at the time at which the choice is being made. However, utility is assumed to consist of two components: the systematic, observable utility which is identical to the conventional microeconomic utility function; and a random term which is intended to capture such effects as variations in perceptions and tastes of individual trip makers [Comenchich and McFadden, 1975]. The relationship is shown in the following formula:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

In which, U_{ij}: random utility of alternative j for individual i

V_{ij}: systematic (observable) portion of utility

ϵ_{ij}: random portion of utility

Further, if the systematic utility V_{ij} is assumed to be a function of the attributes of the alternative j and the characteristics of the individual i, we can get:

$$V_{ij} = b_1 Z_{ij1} + b_2 Z_{ij2} + \ldots + b_n Z_{ijn} = BZ_{ij}$$

In which, b is a row vector of parameters for the independent variables.

$Z_{ij} = f(i, j)$.

Thus, given a set of alternatives C_t, the probability of individual i choosing alternative j from this set of alternatives (P_{ij}) is:

$$P_{ij} = P(U_{ij} >= U_{it}),\ \text{V t ϵ } C_t$$

$$P_{ij} = P(V_{ij} + \varepsilon_{ij} >= V_{it} + \varepsilon_{it}),\ \text{V t ϵ } C_t$$

If we assume that the ϵ's are distributed multinomially normal, the model is known as a probit model. Probit models are computationally cumbersome. If we assume the distribution of the ϵ's is that they are each independently and identically distributed with a Gumbel Type I distribution, the final expression for P_{jj} is the multinomial logit model. The logit model has a tractable, convenient functional form, but it assumes the independence of irrelevant alternatives (IIA), which means the alternatives included in the choice set are independent of each other. When any two alternatives have a non-zero probability of being chosen, the ratio of one probability over the other is unaffected by the presence of absence of any additional alternative in the choice set which often is not the case. To solve the IIA assumption, McFadden (1981) developed a class of models known as generalized extreme value models which includes the MNL and extensions. The nested logit model is one of the more commonly used models in this class. The idea behind a nested logit model is to group alternative outcomes suspected of sharing unobserved effects into nests. Nested logit models are used when alternatives are not independent, or when taste variations exist among individuals in which case require random coefficient models rather than mean-value models as the multinomial logit (MNL).

Route choice models are based on two conceptual steps: identification of the available alternatives (choice set) and choice from a given choice set (specification of systematic utility and functional form). In route choice, alternative feasible paths may be numerous and probably not all perceived by all users [Cascetta et al., 2002]. Random utility models simulate the choice of a decision maker among a set of feasible alternatives and their operational use requires that the analyst is able to correctly specify this choice set for each individual. The assumption of correctly specified choice set may be unrealistic in many practical cases and in particular in modeling route choice when hundreds of paths are potentially available; the result of ignoring this aspect may cause significant mis-specification problems.

Cascetta et al. [2002] summarized different approaches used for the choice-set generation in previous research, as shown in Figure 2.7. The "exhaustive" approach selects all the loop-less paths connecting the origin and destination as the choice set; the "selective" approach only select a subset of topologically feasible paths as the choice set. Selective implicit path enumeration methods are based on rules. One of the best known algorithms is developed by Dial [1971]. The selective explicit path enumeration approach proposes different models related to the available data on paths chosen or perceived by users. The explicit path enumeration methods include two different approaches. In the rule based choice set generation method, perceived paths are obtained as those satisfying some rules. The choice set generation model method first generates a "complete" macro set of paths and then the probability of including a given route in the users' perceived route set is calculated based on specific route attributes.

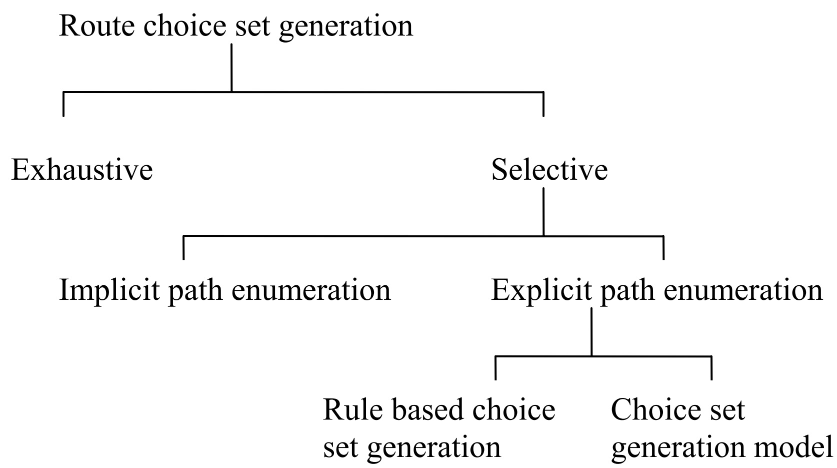


Figure 2.7: Choice Set Generation

Source: Cascetta et al. [2002]

Route Choice and Traffic Assignment

The route choice model that decides which route to take given the origin, destination, and mode of travel of a trip is a central element of the traffic assignment procedure. Traffic assignment is a step of the four-step travel demand forecasting models. In the four-step models, the first step, trip generation, provides the connection between land use and travel. It uses known relationships between trip making and demographics to predict the number of person trips, or 'trip ends', starting and ending in particular geographic areas, or 'traffic analysis zones' (TAZs). The second step, trip distribution, uses characteristics of the transportation network and regional demographics to distribute the trip ends from the generation model to specific origins and destinations amongst the TAZs. The third step, mode split, allocates person and vehicle trips to a particular travel mode. Using Level-of-Service characteristics of each available transportation system, the model ‘chooses’ a mode of travel for each trip based on the relative attractiveness of each competing mode. Traffic assignment is the forth step, in which, the volumes on the transportation system are estimated. The volumes can be present-day volumes on an existing network or forecasted volumes on alternative future systems. Assignment volumes may be expressed as vehicles on a highway network or persons on a transit system. Traditional traffic assignment is performed between traffic analysis zones (TAZs). All trips are assumed to begin and end at the TAZ centroids. It is commonly argued that small displacements of trip ends do not greatly affect route choice, but this assumption needs further validation [Jan et al., 2000].

The basic problem in traffic assignment or transportation network equilibrium modeling is to find the link flows given the origin-destination trip rates, the network, and the link performance functions [Sheffi 1985]. Ortuzar and Willumsen [1990] developed the following classification scheme for traffic assignment (Table 2.2):

Table 2.2: Traffic Assignment Methods

Source: Ortuzar and Willumsen [1990]

		Stochastic effects included?	
		No	Yes
Is capacity restraint included?	No	All-or-nothing	Pure stochastic
	Yes	Wardrop’s equilibrium	Stochastic user equilibrium

All-Or-Nothing

This method assumes that there are no congestion effects, that all drivers consider the same attributes for route choice and that they perceive and weigh them in the same way.

The absence of congestion effects means that link costs are fixed; the assumption that all drivers perceive the same costs means that every driver traveling from i to j must choose the same route.

Wardrop's Equilibrium (User Equilibrium & System Equilibrium)

User-optimal equilibrium theory is developed by Wardrop [1952], who postulated that all used paths between origin and destination require the same travel cost, and an assignment is at equilibrium when no user can improve travel time by unilaterally changing routes. Each individual minimizes or "optimizes" his own travel time or cost. Wardrop's principle implies that all users are assigned to a shortest path between their respective origins and destinations and that travel times and volumes are consistent with each other everywhere on the network. User-optimal equilibrium assignments can be multi-path when two or more paths between an origin and destination have equal travel time. In system-optimal equilibrium, users are assigned in a way to minimize the network wide travel cost. System-optimal equilibrium minimizes the total cost sum of all drivers. It does not occur in practice as some drivers are made worse off. It is used as a benchmark to assess the efficiency of road networks.

Pure Stochastic

Stochastic assignment was developed to relax the assumption of all-or-nothing shortest path assignment used for deterministic user equilibrium model. In this case, either travel times are not known with certainty by trip-makers, or they are imperfect optimizers, or they include factors other than travel time in their decision-making. Stochastic method uses notions of probability theory to distribute trips across several likely paths between origin and destination pair.

Stochastic Equilibrium

Stochastic equilibrium is based on the notion that travelers' knowledge of the traffic situation or the transportation system is imperfect. These procedures recognize that several routes between an origin and a destination might be perceived to have equal travel times or otherwise be equally attractive to a traveler and, as a result, might be equally likely to be used by that traveler. Or, in other words, these procedures treat link costs as random variables that can vary among individuals (given their individual preferences, experiences, and perceptions) rather than deterministically (as is done by the other assignment techniques). Under stochastic equilibrium, all reasonable paths between origin and destination will have flow.

Micro-Simulation

These methods use Monte Carlo simulation to represent the variability in drivers' perceptions of link costs. They rely on the following assumptions: 1) there is a distribution of perceived costs for each link, with the engineering costs (actual cost) as the mean; 2) the distribution is differed in different models, for example, a uniform distribution, or a normal distribution, or other types of distributions; and 3) the distribution of perceived costs is assumed to be independent.

The key to traffic assigning is the underlying behavior assumption of drivers' route choice. The basic premise in assignment is the assumption of a rational traveler, i.e. one choosing the route which offers the least perceived (and anticipated) individual costs [Ortuzar and Willumsen, 1990]. Although a large number of factors are thought to influence route choice, as summarized in the previous section, the generalized cost expression does not incorporate all these elements. The most common approximation of the travel cost considers only two factors in route choice: time and monetary cost; further, monetary cost is often deemed proportional to travel distance [Ortuzar and Willumsen, 1990].

Literature Review Summary

Extensive research has been carried out in the area of route choice. Previous research established theories of route choice decision-making process; and identified route choice factors other than travel time and distance. Most recently, a large number of research efforts devoted to studying the route choice behavior under the influence of traffic information system, the dynamic aspect of the route choice behavior, and the interrelation of route choice, departure time, and trip-chaining decisions.

From the review of the literature, it appears that most of these research results in route choice study were based on stated preference surveys or simulation methods. Few studies were based on revealed preference surveys, and vary little work has been done based on the field observation method. A study based on the real world observations of the actual behavior can help developing a larger body of knowledge in route choice.

Chapter 3

Data Collection

This chapter summarizes the data collection efforts of this dissertation. It starts with a brief summary of the traditional route choice data collection approaches used in the previous transportation studies. The second part is an introduction of the Global Positioning Systems technology used in the data collection of this dissertation. The third part provides an overview of the Commute Atlanta project, the source of the data for this dissertation.

Route Choice Data Collection Approaches

Even though data collection is an important part of the research on route choice behavior, collecting objective link-level route choice data is a very tedious and time-consuming process. For this reason, route choice data are not included in the traditional travel diary data collection methods. In truth, very little empirical work is based on real world observations. Traditional data collection methods for route choice studies include stated preference surveys, revealed preference surveys, and interactive simulations. Many studies used stated preference (SP) data based on hypothetical scenarios; few studies used reveled preference (RP) data. The revealed preference approach analyzes drivers' behavior in the real-life situations based on respondents' reports about previous actions. Revealed preference surveys on behavior during a multi-day period have the ill-effect of survey fatigue and lower accuracy as the survey period extends. The stated preference approach analyzes driver predictions of their behavior in response to hypothetical scenarios. This method may not reflect true route choice decisions under real situations. Interactive laboratory-like experiments that normally involve actual commuters in a simulated traffic system are less expensive and easier to control, but usually restricted to choice between a few alternatives connecting a single O-D pair. Again, laboratory simulations are not likely to reveal the real world decision-making conditions.

GPS technology is now increasingly utilized in transportation research as the technology becomes more accurate and less expensive. Advances in GPS technology make route choice data collection for travel diary studies and other transportation applications a reality. The improvements are both in data quality and data quantity, as well as the addition of new data elements that were once too burdensome or expensive to capture. Recent developments in GIS provide handy tools to manage the large amount of spatial related data captured by GPS units and to post processing to attract route choice information from the raw GPS data [Wolf et al., 1998]. Since the application of GPS technology in route choice research is relatively new, not many established research results are available in this area.

Jan et al. [2000] concluded that GPS is a viable tool to study travelers' route choice patterns. GPS data collection methods can reveal important travel behavior information that was impossible to discern with earlier conventional survey methods. The drawback mentioned in the study is the high cost of the GPS equipment. One limit of their study is the limited sample size. Only around 3000 trips made by 100 households during one week period were available for analysis. They also mentioned since GPS data themselves do not provide information about the underlying reasons why a traveler choose certain routes over others, a post follow-up interview can help to gain insight into travelers' decision-making process. They found that travelers habitually follow the same path for the same trip. However, path deviation increases as origins or destinations become farther apart. Another important finding in this research is that actual travel time tracked by GPS is very close to the calculated network time based on the same path and also to the shortest path time, but the actual travel path is often quite different from the shortest path.

Wolf et al. [2000] conducted a proof-of-concept study to determine if GPS data and a spatially accurate GIS land use database could be used to replace standard travel data collection techniques. The research demonstrates the potential for GPS and GIS in travel surveys, but the study was not capable of identifying related factors of route choice behavior due to the small sample size.

Introduction of Global Positioning Systems

Overview

GPS is the most significant recent advance in navigation and positioning technology. GPS uses satellites and ground equipment to determine position anywhere on Earth. This satellite-based navigation system was launched by the Department of Defense in the 1970s and became fully operational in 1995. Responsibility for the day-to-day management of the GPS program and operation of the system continues to rest with the

Department of Defense, and is carried out primarily by the Air Force [Hofmann et al., 1992].

The initial intent of the GPS system is military applications as well as civilian applications. Although civilian users can use this positioning system at no cost, the GPS system was designed that civilian users would not be able to obtain the same accuracy as the military could. Signal degradation, called Select Availability (SA), was added to the civilian signals. The position accuracy for civilian receivers is 30 to 100 meters with SA enabled. On May 1st, 2000, The US military switched off the GPS signal degradation, random errors decreased from up to one hundred meters to only a few meters for civilian users. When SA is set to zero, the position accuracy of civilian users is increased to 15 to 20 meters. Differential GPS (DGPS) or phase differencing techniques is needed for better accuracies. Modern GPS receivers can improve the original 15 to 20 meter positioning accuracy by their up-to-date electronics and signal processing techniques and microcode to achieve accuracy better than 10 meters (95 percent of the time).

System Components

The GPS system consists of three components: the control segment, the space segment, and the user segment [http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html].

- The Control Segment

The control segment comprises the operational control system which consists of a master control station, worldwide monitor stations, and ground control stations. The main operational tasks of the control segment are: tracking of the satellites for the orbit and clock determination and prediction modeling, time synchronization of the satellites, and upload of the data message to the satellites.

- The Space Segment

The space segment consists of 24 satellites in earth orbit at a nominal height of 20,183 kilometers above the Earth (see Figure 3.1) [B. Hofmann et al., 1992]. To provide a continuous global positioning capability, 24 evenly spaced satellites that are placed in circular 12-hour orbits and are inclined 55 degree to the equatorial plane would provide the desired coverage for the least cost. In any event, this constellation provides a minimum of four satellites in good geometric position 24 hours per day anywhere on the earth.

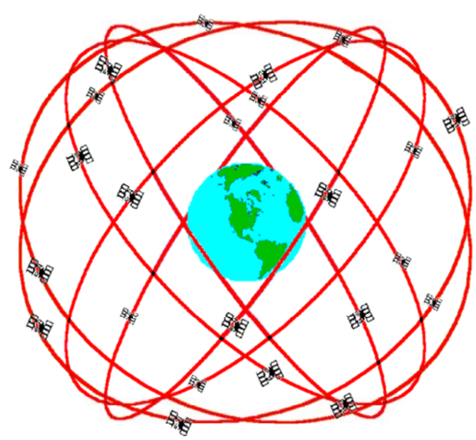


Figure 3.1: GPS Orbiting Satellites

Sources: http://www.mercat.com/QUEST/Intro.htm

- The User Segment

The user segment comprises the receivers that have been designed to decode the signals transmitted from the satellites for the purposes of determining position, velocity or time.

How GPS Works

The GPS system operates on the principle of triangulation; three measurements can locate a point in 3-dimensional space. All GPS satellites synchronize operations so that these repeating signals are transmitted at the same instant. The signals, moving at the speed of light, arrive at a GPS receiver at slightly different times because some satellites are farther away than others. The distance to the GPS satellites can be determined by estimating the amount of time it takes for their signals to reach the receiver. On the satellite side, timing is almost perfect because they have precise atomic clocks on board, but the clock on the receiver side is usually not that accurate. Since both the satellite and the receiver need to be able to precisely synchronize their pseudo-random codes to make the system work, one more satellites are needed to compromise the time error of the GPS receiver. When the receiver estimates the distance to at least four GPS satellites, it can calculate accurate position in three dimensions include latitude, longitude and altitude (see Figure 3.2).

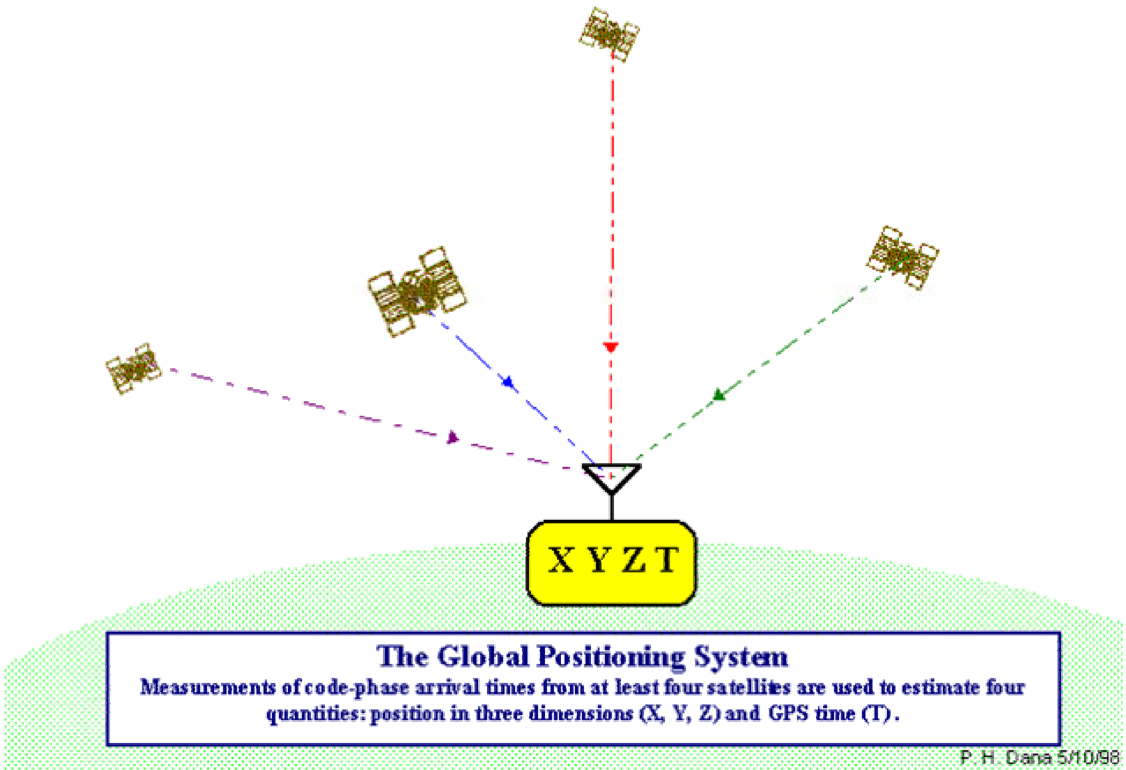


Figure 3.2: GPS Position Calculation

Source: http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html

Sources of GPS Data Error

There are several factors associated with the GPS system that can cause errors in GPS position information [Parkinson and James, 1996].

- Atmosphere-induced errors

Atmosphere-induced errors are caused due to the fact that GPS signals do not travel at the vacuum speed of light as they transit the atmosphere. Now that SA has been removed, one of the most significant error sources is ionospheric error. The ionosphere is the layer of the atmosphere ranging in altitude from 50 to 500 km. It consists largely of ionized particles which can exert a perturbing effect on GPS signals. Ionospheric errors can range from 5 to 7 meters. The troposphere is the lower part of the earth's atmosphere that encompasses the weather. It is full of water vapor and varies in temperature and pressure. The troposphere layer cause relatively less error range from 0.5 to 0.7 meters. The most effective way to handle these errors is to compare the relative speeds of two different signals, normally called the dual frequency correction. This correction is sophisticated and is only possible with advanced receivers.

- Satellite orbit errors

GPS Satellite orbits can vary slightly over time and require periodic adjustment by the system maintainers. Since the orbits vary, errors can exist in the satellite ephemeris (location) data used in triangulation calculations. The range of the orbit errors is normally less than 1 meter.

- Satellite clock errors

The atomic clocks used in the GPS satellites are very precise but not perfect. Minute discrepancies can occur, and these translate into travel time measurement errors. The atomic clock error in a day is about 3.5 meters.

- GPS receiver errors

Receivers may introduce some errors by themselves in measuring code or carrier. In high quality receivers, however, these errors are negligible (less than one millimeter) for carrier phase and a few centimeters for code phase.

- Multi-path errors

Multi-path is caused by reflected signals from surfaces near the receiver such as buildings or cars that can either interfere with or be mistaken for the signal that follows the straight line path from the satellite. Multi-path is difficult to detect and sometime hard to avoid.

Civilian GPS technology can be supplemented with other correction technologies in order to provide an accurate location at all times. Dead Reckoning is one of the several methods used to provide corrections. Dead Reckoning is used for applications that need continuous positioning, even in places where GPS signal is unavailable, such as tunnels, parking garages, and urban canyons. Dead Reckoning uses extra sensors installed in the vehicle to measure vehicle speed and direction. By combining this information with the GPS data, it can determine the vehicle's current position based on the last known position and travel speed and direction, even when GPS signals are blocked or reflected. The integrated GPS/Dead Reckoning system provides a much more robust vehicle position.

Applications in Transportation

GPS has drastically changed methods of navigation and is becoming important in everyday life. As the GPS receivers have been miniaturized to just a few integrated circuits and are becoming as cheap as couple hundred dollars, the technology is accessible to virtually everyone. These days GPS is finding its way into cars, boats, planes, construction equipment, movie making gear, farm machinery, even laptop computers.

GPS technology is now increasingly utilized in transportation research as the technology becomes more accurate and less expensive [Wolf, 2000]. Some examples of the GPS applications in transportation research include, but not limited to the following:

- Control points in survey applications
- Spatial accuracy check of the GIS database
- Vehicle navigation and automatic vehicle location
- Travel time and traffic system performance studies
- Travel behavior surveys

Previous GPS Travel Behavior Studies

To improve the data quantity and quality of travel survey, transportation researchers have carried out pilot studies that use GPS technology to supplement the traditional data elements collected in paper or electronic travel diaries. A proof-of-concept study was also carried out to examine the possibility of completely replacing the traditional travel diaries use GPS data collection method [Wolf, 2000].

GPS technology can be used either in an electronic travel diary, which is a carry-on-package turned on and off by the traveler, or in a passive in-vehicle data recorder which is turned on and off automatically together with the vehicle engine. The former system has more flexibility since it is carried by the traveler and can record trips made by travel modes other than personal vehicle such as walk and transit. On the other hand, more errors are introduced due to the fact that travelers may forget to turn on or off the unit in time. The latter system is installed in a certain vehicle and can only record trips made by that vehicle, but it is more accurate and reliable since no human interference is needed.

Wolf [2000] summarized four important studies have used GPS to supplement travel behavior data collection. Two of these studies, conducted in Lexington and the Netherlands, explored the use of handheld electronic data collectors. The other two studies, conducted in Austin and Quebec City, used the passive in-vehicle GPS systems. Table 3.1 summarized the primary characteristics of each study. Although factoring in all data losses, a total of 63 percent of all PDA-recorded trips contained good GPS data for further analysis in the Lexington study. The study concluded that GPS can be used to improve travel behavior data collection. The Netherlands study revealed both problems and promise in GPS application of travel behavior study. The study found that GPS units are possible to monitor all travel modes, but data completeness and quality varies by mode. The Austin study incurred significant amount of data loss and difficulty in trip ends identification. The Quebec study revealed several problems with the equipment, including GPS acquisition times, power supply stability, and data storage limits. Even though, the researchers concluded that it is possible to collect multi-day data using passive in-vehicle GPS units. In conclude, all the projects mentioned above confirmed the capabilities of GPS technology in aspect of improving travel study, but challenges remain in aspect of better application design that can avoid data loss and improve data accuracy, as well as post process that can attract trip level information in an efficient manner.

Current Large Scale GPS Based Transportation Studies

While previous GPS studies tended to be in small sample size and short durations, several recent studies have larger sample sizes and longer durations. One of the recent studies is part of the Swedish Intelligent Speed Adaptations (ISA) study, which installed GPS based unites in around 300 vehicles in three Swedish cities and observed the vehicles for up to two years [Axhausen et al., 2003]. This study is focused on the traffic safety effects of in-car speed information systems. In another study, Doherty et al. [2004] combined GPS and GIS technologies with a recently developed computerized activity scheduling survey. The developed system has the potential to simultaneously observe detailed spatial-temporal activity-travel patterns and underlying decision processes of individuals within a household over long periods of time, while at the same time minimizing respondent burden.

Table 3.1: Summary of Four Previous GPS Based Travel Behavior Studies

Source: Wolf [2000]

	Electronic Travel Diary with GPS		Passive In-vehicle GPS System	
Study location	**Lexington, KY**	**The Netherlands**	**Austin, Texas**	**Quebec City, Canada**
Survey period	6 days	4 days	1 day	1 - 3 weeks
Time frame	Fall 1996	Winter 1998 – Spring 1999	1997-1998	December 1998 – March 1999
Purpose	To test integration of GPS with self-reported travel behavior; to compare differences between two methods	To demonstrate the success of using a handheld PC with GPS for travel data collection for all modes	To compare GPS data with reported trips and start/end times; to compare trip routes/times with models; to examine speed profiles	To test feasibility of recording personal travel with GPS over multi-week periods
Sample size	100 households • 1 vehicle per household	151 people	203 households • up to 3 vehicles / household • total of 356 vehicles	4 researchers
Modes collected	Vehicle only	All modes	Vehicle only	Vehicle only
Additional data collection methods	Recall interview for 1 day of survey	Paper diaries were provide to record trips not captured by GPS equipement	Paper diaries during survey CATI retrieval	Trip purpose and passenger load recorded on laptop or paper
GPS interval	3 seconds	4 to 10 seconds (by mode)	1 second	4 or 5 seconds (by receiver)
Differentially corrected	No	Yes Real-time, FM	Yes post-processed	Yes Real-time, beacon
Routing analyses	Yes, Univ. of WI has compared path variations between matched trips	Limited	Not yet	Yes, algorithms developed for map matching and stop detection
Positive findings	CASI with GPS can improve travel survey data quality	It is possible to capture GPS data for all travel modes	GPS can be used to identify missed trips	It is feasible to collect multi-day in-vehicle GPS travel data
Findings of concern	User activation issue GPS power supply loss Data logging frequency Difficulty in matching recall, PDA, GPS trips GPS acquisition time	GPS and DGPS reception varies by mode Respondents did not use equipment for specific trip modes and purposes Lost data for 22 of 151 people	PP storage requirements Lost base station data Lost vehicle data GPS accuracy GPS acquisition time Trip end identification	GPS acquisition time Power supply stability Data storage limits Cold temperatures: LCD screen

Commute Atlanta Project

The data set used in this dissertation is taken from The Georgia Institute of Technology Commute Atlanta project. The Commute Atlanta project is an instrumented vehicle research program funded by the Federal Highway Administrations (FHWA) Value Pricing Program, the Georgia Department of Transportation and Georgia Tech. The passive in-vehicle data collection unit designed for the project provides accurate second-by-second time, speed and location information and also an accurate itinerary of vehicle trips, including those short, intermediate, and infrequent stops that would otherwise be missed in traditional travel diary data collection methods. The project has deployed instrumentation in 487 vehicles from 268 representative households in the 13-county Atlanta metro area (Figure 3.3) and has collected second-by-second speed and position data for more than 600,000 trips during the first ten months of the data collection. The project includes household interviews that establish household and driver demographic and socioeconomic activities and employer surveys examine commute options and parking policies.

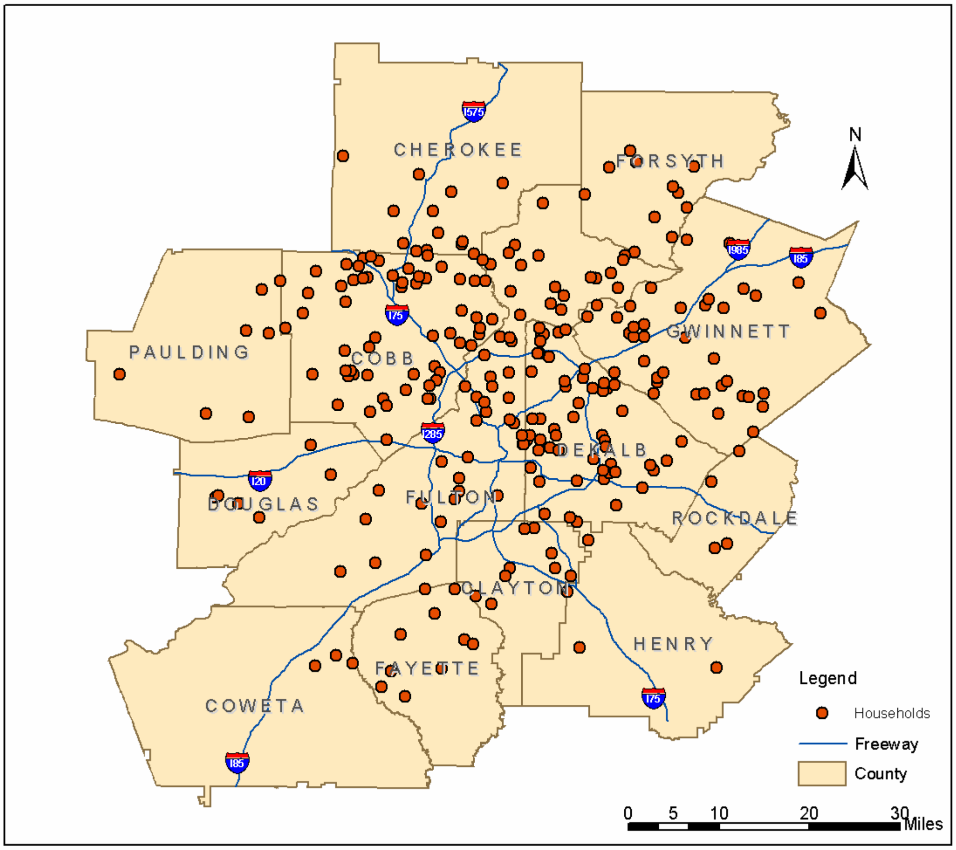


Figure 3.3: Commute Atlanta Project Participating Household Distribution

Data Collection Equipment

The in-vehicle Event Data Recorder (EDR) used in the study is shown in Figure 3.4 and Figure 3.5. Figure 3.4 shows the outside look of the EDR. Figure 3.5 shows the EDR and accessories. Major components of the EDR include CPU, power System, cellular transceiver, GPS, and other Sensors. The optional connections connecting to the EDR include 6 on/off sensors, 2 serial connections for OBD and an extra input. These sensors can detect seatbelt usage, ODB data, braking, windshield wiper etc. The digital cellular transceiver is capable of sending data through low cost short message service (SMS) or larger volume circuit switched data. The Commute Atlanta project data collection system map is shown in Figure 3.6.

Figure 3.4: Outside Look of the EDR

Figure 3.5: EDR and Accessories

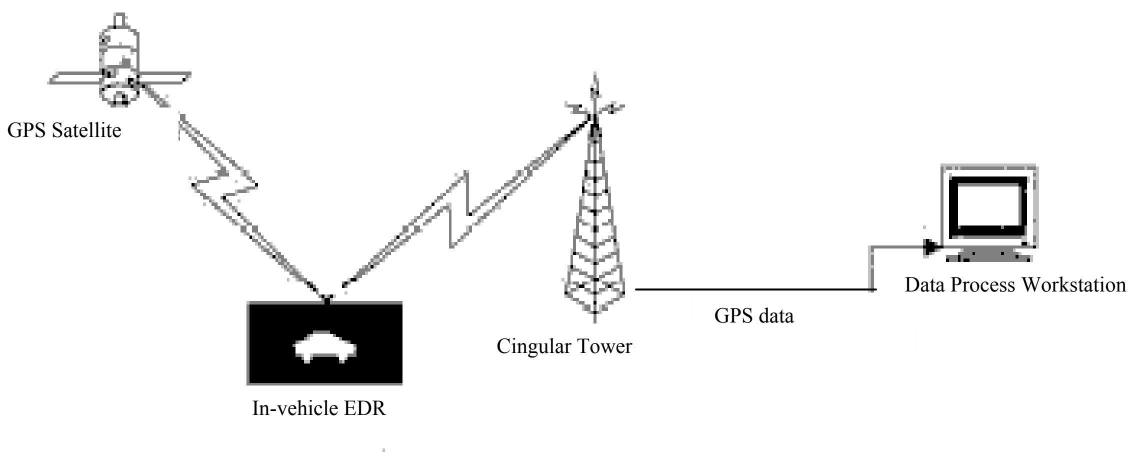


Figure 3.6: Commute Atlanta Data Collection System Map

The EDR turns on and off automatically with the vehicle ignition and no human input is required. Recorded data are downloaded automatically over a cellular connection every week. These features make the EDR a practical option to monitor travel behavior 24 hours a day during multi-day period. Based on the product manual, the GPS receiver used in the study has 12 parallel-channels. It has a position accuracy of 10 meters and is capable to acquire satellite signal within 8 seconds under hot engine start situation and 45 seconds under cold engine start situation. Since the vendor's specifications are usually based on a desirable number and geometric distribution of in-view satellites which is often not the case in the real world, the actual quality of the GPS data acquired in the study varies from time to time. It usually takes up to 15 seconds in hot engine start condition and 60 seconds, sometimes as long as 2 minutes, in cold engine start condition for the unit to function at full accuracy. Therefore, data quality can be unstable within first 2 minutes after the trip start. Under the worst case scenario, invalid GPS positions scatter within the first 1 to 2 minutes of trip start (Figure 3.7).

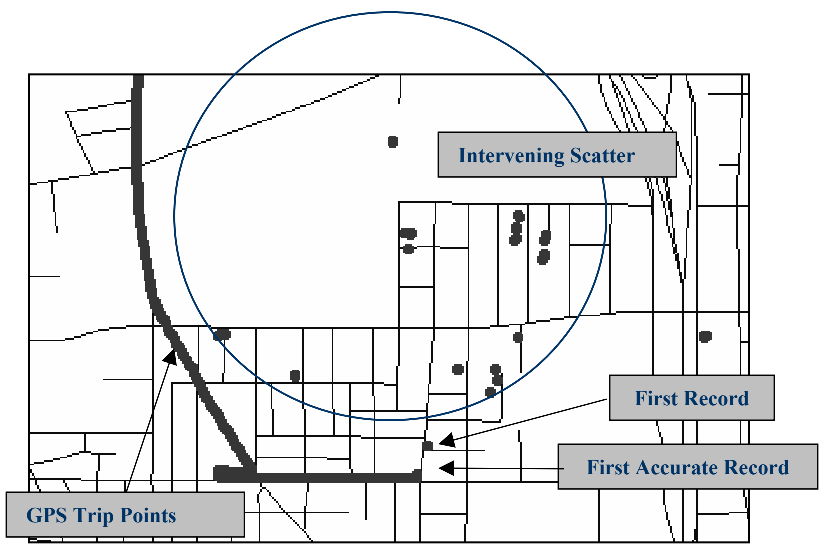


Figure 3.7: Scattered GPS Points at Trip Start

Big cities pose more problems for GPS data collection. Tall buildings often block GPS Satellite signals, or the signals are reflected off of the buildings causing multi-path errors.

Both these problems cause the vehicle's positions appear to jump around when plotted on a map (Figure 3.8). Although the integrated GPS/Dead Reckoning system can provide more robust vehicle position under these conditions, the combined system introduces other types of error. For example, a test of the combined system of GPS and Dead Reckoning showed the unit provided rotated location when the vehicle starts in a backward motion such as backing out a parking space. Hence, it is not included in the GPS unit used in the Commute Atlanta project.

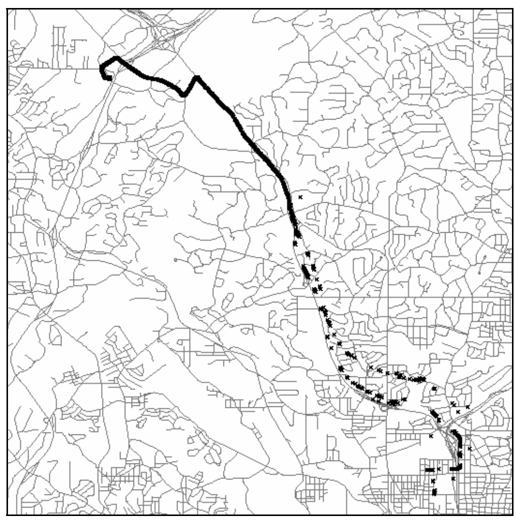
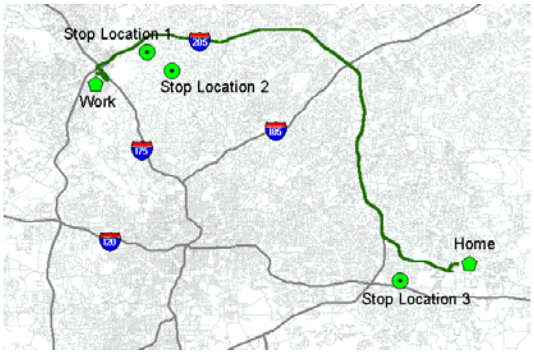


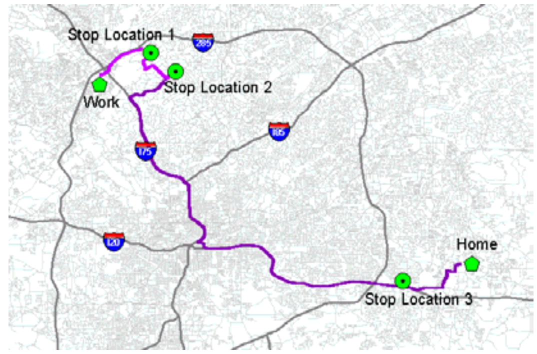
Figure 3.8: A Scattered Trip Caused by Invalid GPS Location Acquisition

Commute Activity Examples

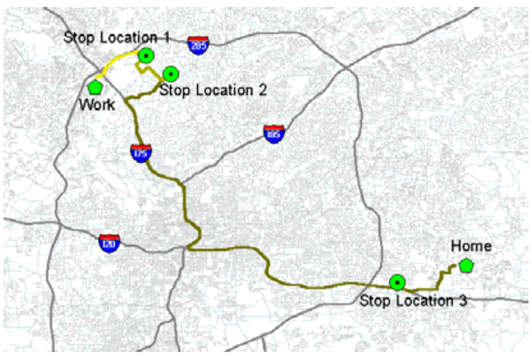
The data collectors instrumented in the Commute Atlanta project generate a very rich set of vehicle activity data for further research analysis. A visual example of a five-day commute, shown in Figure 3.9, illustrates the level-of-detail of the information available from the data set. Figure 3.10 summarizes the rich information contained in the sample commutes of Figure 3.9. In Figure 3.10, sections in different colors from the same commute day are trip segments of the commute journey. Engine-off and engine-on stops are represented by stars.



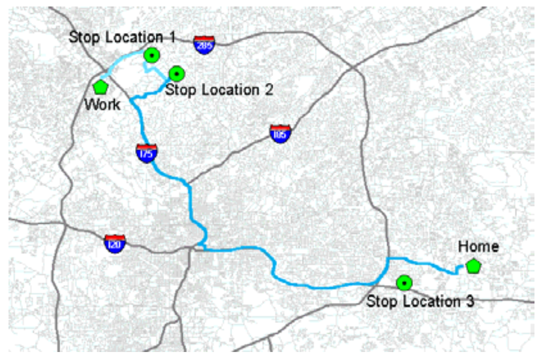
Commute Day One



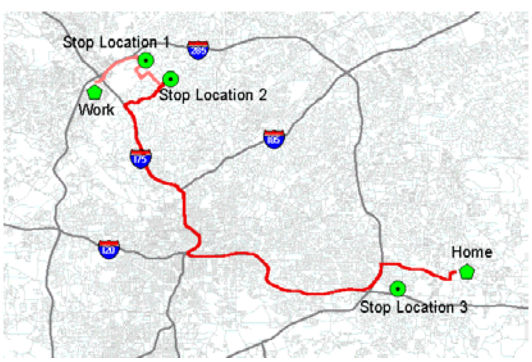
Commute Day Two



Commute Day Three



Commute Day Four



Commute Day Five

Figure 3.9: A Five-Day Commute Journey Example

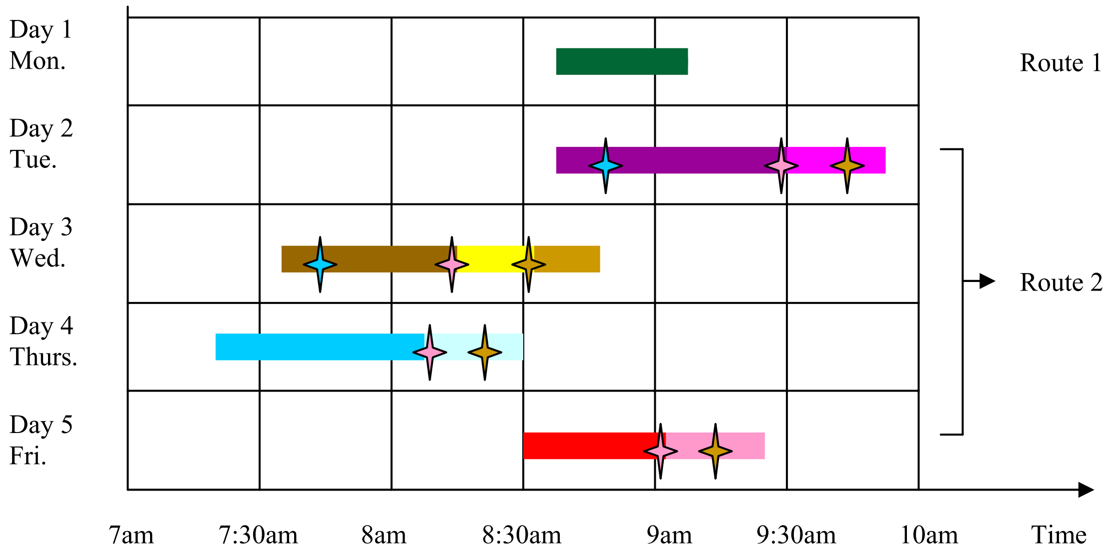


Figure 3.10: Commute Information Summary

Chapter 4

Route Choice Data Generation

Placing the GPS data directly onto a GIS based digital map may produce results in which the GPS points and the GIS road links are not congruent, because both the GPS location data and the GIS road network are not one hundred percent accurate. Due to the fact that vehicles almost always travel on the road network, the methodology that translates the GPS measured position onto the digital road network is called map-matching. The algorithm needs to reconcile two types of error, the inaccurate GPS position and the inaccurate digital road network, and associate the position of a traveler in the real world with a position on a digital road link. Through map-matching algorithms, a route in the format of a sequence of arcs between trip origin and destination can be generated from a sequence of GPS positions.

GIS Data Accuracy

The accuracy of the digital road network is crucial to the success of GPS data map-matching and interpreting results. Depending on the data source and generating method, the accuracy varies greatly (Figure 4.1). The Tiger (topographically integrated geographic encoding and referencing) files of the census bureau were created for census-tracking purposes. They are readily available, inexpensive, and widely used, but have low spatial accuracy. Error in the Tiger files can be as large as 30 to 50 meters.

Commercially available maps are more accurate for metropolitan areas, but are also more expensive.

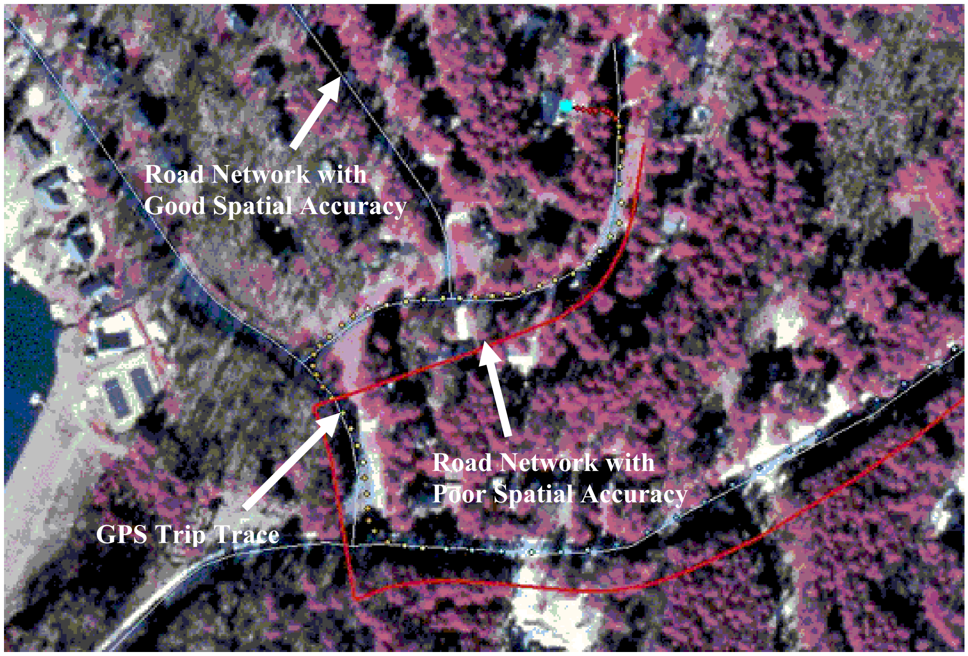


Figure 4.1: Two GIS Road Networks at Different Spatial Accuracy

Literature Review on Map-matching Algorithms

Most of the development in map-matching algorithms takes place in the area of in-vehicle navigation systems. In these systems, vehicle position, speed and heading information are captured, and map matched in real time. The map-matching algorithm must identify the road network link being traveled using only the current and historical information.

White et al. [2000] discussed the basic ideas of point-to-point, point-to-curve, and curve-to-curve matching along with the advantages and disadvantages of each method. Their study implemented and tested four different algorithms. The first algorithm finds the closest arc to the GPS tick and projects the point onto that arc. The second algorithm uses “heading” information in addition to the first algorithm. Since both algorithms do not use “historical” information, both of them are unstable. The third algorithm uses topology (road network connectivity) information. The problem with this algorithm is that one bad match can lead to a sequence of bad matches. The forth algorithm uses curve-to-curve matching by selecting the closest road curve to the GPS points curve and projecting the point onto that road curve. The algorithms were tested on four in-town routes and the best algorithm correctly matched between 66 percent and 86 percent of the GPS points. Algorithm 1 had the worst performance, algorithm 2 performed reasonably well, and algorithm 3 and 4 did not out-perform algorithm 2 which is considerably simpler. Incorrect matches were more likely to occur when the road segment is relatively short.

Joshua [2002] first discussed the main shortcoming of the algorithms using only geometric information. The author presented a weighted topologically-based matching algorithm that generated a line passing through the continuous GPS points, and evaluated proximity and orientation with the street network. He then measured proximity of the GPS point to the road line, similarity in the orientations between the GPS line and the road line, and the intersection between the GPS line and the road line were developed. Based on those measurements, he calculated a weighted score used to evaluate the correct match.

Yim et al. [2002] described a road-mapping algorithm in which data points are matched to an underlying network map and a set of possible paths is built through the matching network links. For each data point the software identifies all the possible links located within a specified accuracy distance. This set of links represents the possible current positions of the probe vehicle, and is added to the sets of links generated by earlier data points. Each of the new links is examined to determine whether it can be reached from any of the previous possible positions. As successive data points are added, the number of possible paths changes until eventually the set of possible paths is reduced to zero or one. If the set of paths is reduced to one, the actual path travel has been determined.

Makimura et al. [2002] used a five-step process to match the GPS data in a car navigation system with the GIS data. Basically, this method is a shortest path method. First, they used the direction data to separate the continuous car navigation data into origins and destinations i. e., the points where the direction changes suddenly. Next, they joined these data with a straight line and extracted the GIS road data included within a radius of 30 meters from these joined route data. Finally, they used these GIS road data to search for the shortest route and determined the route.

More complex methods are generated in other researches from Lamb and Thiebaux [1999], Scott [1994] and Kim and Kim [2001]. These methods used fuzzy technology, Kalman filters, or Markov models to provide estimates of the vehicle location on each of the hypothesized road segments onto the road network. These conditional probability based approaches differ from conventional geometric based schemes by not performing any explicit map-matching step. Although, approaches based on conditional probability such as Kalman filtering may be more robust and can recover from false positioning quickly, their implementation is more complex. It is hard to tell for sure which type of method is better since no standard data sets exist for the evaluation of different map-matching algorithms and very few of such evaluations have even been reported.

Route Generation Algorithm[1]

GPS trip data are updated on a second-by-second basis. Although the data provide very detailed information, they also raise challenges on data process procedures. For example, at one-second interval, one vehicle can generate 3,600 data points of vehicle activity per hour. For large scale studies such as the Commute Atlanta study, the data size reaches tens of gigabytes. The map-matching algorithms and the overall data process procedures have to be highly automated in order to handle the large volume of data in an efficient manner. In this study, the map-matching process is implemented in post process instead of real-time, as in the navigation systems. Both the trip origin and destination information are known before the map-matching step. Hence, a topology based method is more appropriate and easy to implement. More specifically, map-matching implemented in the study is based on the shortest path function in ArcInfo™. The generated route for each trip is in the format of a sequence of road network links. This sequence of road network links is an accurate representation of the route traveled and can be used in route choice studies.

[1] The specific map matching program and data processing routines as well as instrumented vehicle data formats and scripts employed in this research are subject to a confidentiality agreement and licensing provisions. As such, only general descriptions of these methods are provided in this dissertation. For additional information, contact Dr. George Harker, Director of the Georgia Institute of Technology Office of Technology Licensing.

Perl script language and ArcInfo™ Workstation AML macro language were chosen as the data process tools. Perl is an interpreted programming language known for its strong text-manipulation functions and also its flexibility. ArcInfo™ workstation has the network analysis functions most suitable for the task of map-matching.

Data Process Steps

The data process is shown in Figure 4.2.

Step One:
Organize trip data of the same driver by time sequence and add previous trip's end position as current trip's start position

↓

Step Two:
Generate ArcInfo™ coverage file for each GPS trip file and label each GPS point based on data quality

↓

Step Three:
Screen out trips that completely or partially fall out of the extent of the study area. Screen out trips that are not real trips

↓

Step Four:

- Buffer around the GPS coverage to generate a small road network
- Generate ArcInfo™ coverage for each stops file of the GPS trip
- Run the shortest path function based on the stops file and the small road network to generate the traveled route
- Spatial join the generated route with the original GPS trip

Figure 4.2: Route Choice Data Process Flow Chart

The process starts with a batch of GPS ASCII files downloaded from the cellular network and contain individual trips made by several drivers during a certain time period. Each file represents a trip and contains vehicle position, speed and heading information. A trip is started at the time when the vehicle engine starts and ends at the time when the vehicle engine stops. This batch of GPS data need to run through several data preparation steps before the map-matching step.

The first step of the data processing sorts the GPS ASCII files that represent a sequence of trips made by the same vehicle in the order of trip start time. To get accurate trip origin and destination positions, the destination position of the previous trip is used as the origin position of the current trip since the vehicle is not supposed to move while the engine is off. This step is necessary because the trip origin position is normally not accurate in case of signal acquisition delay. On the contrary, the trip destination position is accurate since the GPS unit generally has achieved full accuracy by the time a trip ends. The origin position of each driver's first trip in the batch is retained.

The second step generates the ArcInfo™ coverage file (the native format of the GIS file used in ArcInfo™ Workstation) for each trip based on the GPS ASCII file. Because different sources can cause potential error in the GPS data, not all the GPS data points are accurate. The data cleaning procedure verifies the quality of each GPS point based on the number of available satellites and the Position Dilution of Precision (PDOP) values, a measure of the current satellite geometry, determines whether a certain point is accurate enough for further processing steps.

The third step further screens the data at the trip level. In this step, trips do not meet map-matching requirements are screened out. A trip should be within the study area where the GIS road network covers so that the trip points can be matched to the underlying road network. A point-in-polygon overlay is performed between each trip coverage file and the study area polygon. Trips that completely or partially fall out of the study area are screened out. GPS files that are not real trips, but only engine turned on-and-off without moving, are also screened out.

The fourth step is map-matching. A visual description of the map-matching process is shown in Figure 4.3. In the map-matching process, the shortest path function in ArcInfo^{TM} determines the minimum-cost path to reach a series of intermediate stop locations in the user-specified order. These intermediate stops are not real stops made by the driver. Instead, they are generated by sampling the GPS trip points every minute and whenever the driver makes a turn. The path created is based on the link impedances encountered in the network. In this study, link impedance is the link distance. The shortest path function runs on a small network that consists of links fall within a 100-meter buffer of that specific trip. The purpose of the process is to find the shortest path through the small network that connects all the intermediate sampling GPS points. The algorithm starts at the node designated as the trip destination point, and evaluates the impedances encountered reaching the next sampling GPS point. Potential least-cost paths are evaluated, and a path with the least cumulative impedance is constructed. This process is repeated traversing from the second intermediate sampling point to the third one and so on until all the points have been visited.

Original GPS points | Small network around GPS points

Intermediate sampling GPS points | Route identified

Figure4.3: Map-Matching Steps

The road network used in this study is the Georgia Department of Transportation Digital Linear Graph (DLG) road database. This dataset provides a 1:2,000,000-scale road layer with full topological structuring. The digital network is organized into links and nodes. A link is a representation of the road segment, and a node is placed where different road segments connect with each other. Topology information of the road network includes the connectivity, adjacency and proximity characteristics of the network. Yet, the accuracy of this data has not been strictly tested. Overlay of the road network with digital photo and GPS trip data shows that the accuracy of the road network is actually lower than the GPS data, but both the GPS data and GIS network are overall accurate and overlay well with the DOQQ photos.

The output of the shortest path function is written as a route subclass of the network coverage. The last step of the data process spatially joins the route with the original GPS coverage, and assigns the correct network link information together with the associated RC attributes to each GPS point.

Algorithm Performance

A total of 17,486 trips made by 110 vehicles were used to test the performance of the map-matching procedure. After screening out trips that have less than 3 valid GPS points, short trips that sum of non-zero speed values in the trip is less than 60 mile per hour, the remaining trips went through the map-matching algorithm. Among them, routes were generated successfully for 93 percent of these trips; the remaining 7 percent of the trips failed the process.

Seven percent of the trips failed the map-matching process due to the error in the underlying GIS base map, including topological errors, missing links, and inaccurate link configurations. First, road links that are continuous in the real world should be topologically connected (Figure 4.4). For example, a node should exist at the intersection of two crossing streets so that shortest path function can trace the GPS points from one street to the other. On the other hand, at a location where a local road crosses over a freeway, a node should no exist since it is impossible to turn from the local road onto the freeway directly in the real world situation. Second, in large metro areas, new road constructions are taking place everyday, but the GIS data are only updated quarterly or yearly. This can cause missing road links in the GIS road database (Figure 4.5). Road widening or other types of road constructions can change the road configuration. The GIS database is hard to catch up with those changes (Figure 4.6). In case of multiple-lane freeways, single linear feature in the GIS database can hardly represent the freeway section that consists of up to 12 lanes and is as wide as 150 feet total. Hence, on freeways with multiple lanes, gap between GPS points and road center line can be larger than 75 feet (Figure 4.7). Off-road-network driving can also cause problem in map-matching, for example, a cut-through using parking lot or vacant property makes the vehicle run off the road network and causes the map-matching algorithms fail (Figure 4.8).

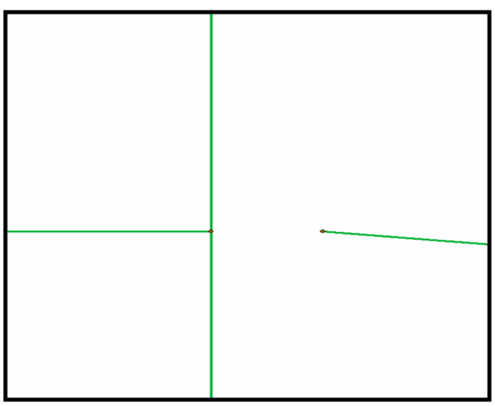


Figure 4.4: A Discontinued Road Link

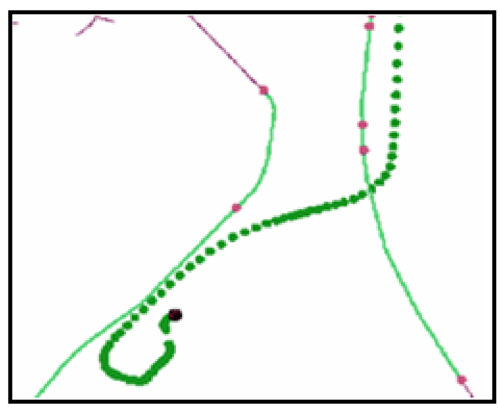


Figure 4.5: A Missing Road Link

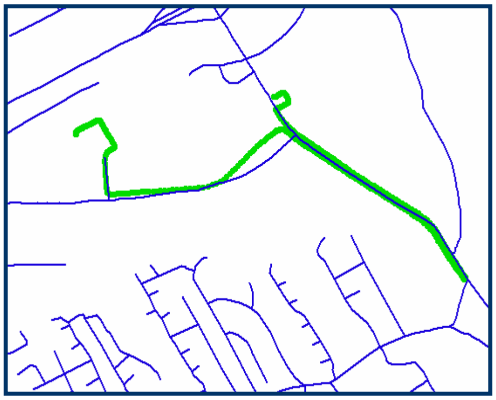


Figure 4.6: Road Configuration Change

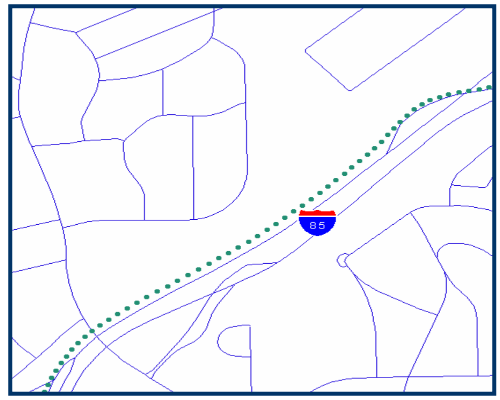


Figure 4.7: Freeway Shape Accuracy

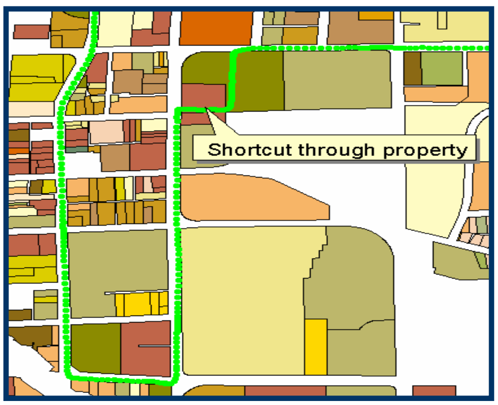


Figure 4.8: Off Road Network Shortcut

Chapter 5

Derive Commute Level Information

The GPS unit records vehicle activities at one-second interval, a vehicle that runs one hour per day generates 3,600 activity records. Deriving commute level information from the huge set of raw data presents a challenge. For example, the network of 100 vehicles equipped with the passive in-vehicle GPS units log more than 2.5 million vehicle-second of activity during one week period. This chapter addresses the methodologies to derive commute level information, including commute start and end time, origin and destination locations, travel time and distance, and trip itineraries. The emphasis of the chapter is to explore methods that can handle huge dataset and can derive correct information with little or no human interference. This chapter also addresses issues that need to be considered in order to get accurate trip level information from the GPS data set.

Identify Possible Morning Commute Journey

A series of trips with the first trip starting at home, the last trip ending at the work place, and all trips intermediate, that take place during the morning commute time-period on a given day are considered a single morning journey-to-work. This dissertation develops a series of procedures to differentiate the morning commute activities from the other vehicle activities during the day.

Screen Based on Date and Time

The morning commute time period is currently defined as 5 a.m. to 10 a.m. local time Monday through Friday. Vehicle activities that occurred on public holidays are excluded from the dataset. Since the original GPS data's date and time are in Greenwich Mean Time (GMT) format, date and time are first translated from GMT time to Atlanta local time. GMT time minus 5 hours is Atlanta local time (US & Canada Eastern Time). GMT time minus 4 hours is Atlanta local Daylight Saving Time. Daylight saving time begins each year at 2 a.m. on the first Sunday of April. Standard Time begins each year at 2 a.m. on the last Sunday of October.

Screen Based on Home and Work Locations

Due to the signal acquisition delay in cold engine conditions mentioned in the previous chapter (it normally takes the GPS unit up to 60 seconds to start acquiring valid position information in the cold-engine start condition), the trip starting position provided by the GPS unit may have an error. Since this study records all the vehicle activities during the study period, and trips take place sequentially, the previous trip's last known position is used as the current trip's starting position because the vehicle is not supposed to move if the engine is turned off. The trip's ending position is mostly accurate since the GPS is fully functional when a trip ends.

The household travel survey has the home address of each household and the work address of each worker in the household geo-coded in Latitude and Longitude format. However because the geo-coded household survey address data are incomplete and inaccurate, among the 487 vehicles from the 268 household in our study, only 223 vehicles have work addresses and 330 vehicles have home addresses. Even if we define trips with the starting positions that fall within a 1000 feet buffer of the home locations are considered starting at home, and trips with the ending positions that fall within the 1000 feet buffer of the workplace locations are considered ending at the workplace, commute journeys could be identified only for 72 vehicles. To solve this problem, this dissertation developed a script that identifies driver's home and work locations based on the driver's activity pattern from GPS data set itself.

The script works as follows:

- Based on the fact that the first trip of a day usually starts at home, starting positions of the first trips during a day are candidates of the driver's home location. Based on the distance between each pair of these location candidates, two locations are assumed identical if the distance is within 1000 ft; other wise, the two locations are assumed distinctive. A location that occurs most often is determined as the home location of that driver.

- Similarly, ending positions of the last trips during the commute time period defined are candidates for the driver's work location. Based on the distance between each pair of these location candidates, two locations are assumed identical if the distance is within 1000 ft; other wise, the two locations are assumed distinctive. A location that occurs most often is determined as the work location of that driver.

Using this method, morning commute journeys of 214 drivers can be clearly identified based on 3 months worth of vehicle activity data. The author examined the commute activities of the 72 vehicles that were identified by both the script and the geo-coding method. The result shows that the home and work locations generated by the scripts correspond to those generated by the geo-coding address method, except one commuter who works in a hospital and commutes work-to-home during the morning period. That commuter was excluded from the sample. Among the remaining 213 vehicles, 11 vehicles were dropped from the set because they had fewer than 10 commutes. Other than those vehicles, manual examination identified that one vehicle actually traveled to two destinations that are close to each other. Eight vehicles were dropped because they contained bad trip files due to GPS error.

Screen Based on Household Survey

Based on the household survey, three vehicles were deleted from the set because those vehicles were shared by drivers already included in the sample. Eight drivers were excluded from further analysis because they are either homemakers or retirees based on the occupation information in the household survey. Even though these drivers traveled daily from home to a certain location during the morning commute time period, it is not clear whether those trips are actually commute trips. The remaining 182 drivers from 138 households work full time or part time at a fixed working location and do not share the vehicle with other household members.

Identify Intermediate Trip-Chaining Stops

As the drivers may or may not turn off the engine when they stop, trip itineraries recorded by the GPS units sometimes do not represent the actual trip itineraries made by the driver. For example, one GPS trip file can include one or more drop-off or pick-up trips in which the driver did not turn off the engine (Figure 5.1).

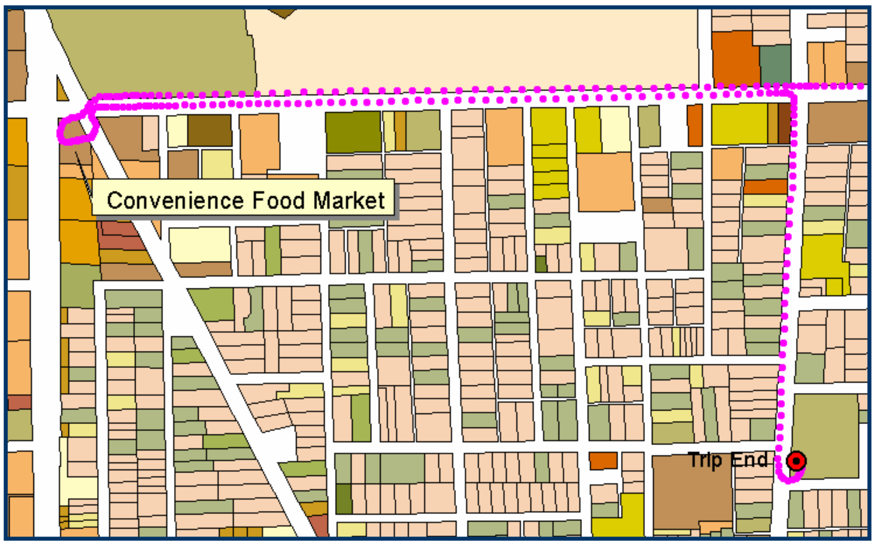


Figure 5.1: A Drop-Off Example

Trip-chaining stops made during the morning commute have been divided into two types: Engine-off stops take place when the driver turns off the engine during the stop. Engine-on stops take place when the driver does not turn off the engine during the stop. Identification of the engine-off stops is straight forward. On the contrary, it is impossible to identify the engine-on stops without a close examination of the actual trip data. A separate script was developed to identify this type of stops based on the following criteria:

- The section of trip points during an engine-on trip-chaining stop should include points fall out of a 75-foot road network buffer.

Based on the accuracy of the GPS and GIS data, the distance between a certain GPS point and the GIS road network is almost always within 75 feet if the driver is actually traveling on a road segment. If the trip points fall out of the 75 feet buffer, possibly it is because the driver actually leaves the road network and enters into a parking lot.

- Stop duration should be longer than 2 minutes.

Lower duration criteria may result in false engine-on stop detection or detect a single stopping behavior multiple times. On the other hand, higher duration criteria can miss stops with short duration. Duration criteria including 30 seconds, 1 minute, 2 minutes and 3 minutes were tested to find the optimum criteria based on a set of test files. The 2 minutes duration criterion is the most reasonable threshold balancing the desire to identify all real stops against the unnecessary detection of false stops. Wolf [2000] also concluded the same duration criteria in her study.

- Speed should be less than 30 mile per hour when the stop begins.

This criterion is set to avoid false stop identification when a driver is traveling along a freeway section. Due to the lower GIS data accuracy on some freeway sections, trip points may fall out of the 75 feet buffer on some freeway sections occasionally even when the driver actually does not leave the road network. This is because the freeways are wide and the road centerline is not an accurate representation of all the lanes on freeway sections. The criteria should not impact trips on local streets because the speed is usually less than 30 mph when a driver makes a turn and leaves the road network.

- The stop duration includes at least one point that has a speed value larger than 5 mph.

Every time the driver turns on the vehicle engine, a GPS file will be recorded. A false trip will be recorded if the driver actually turns on-and-off the engine without traveling any distance. This criterion is used to avoid the situation described above, and make sure the trip is actually a real trip.

- Assume no two trip-chaining stops will take place within a five minutes time period in order to avoid counting one stop multiple times.

Even during the same trip-chaining stop, some of the trip points will fall within the 75 feet buffer of the road network occasionally. If based on the previous criteria only, the algorithm will assume that the vehicle gets back to the road network and makes multiple trip-chaining stops, which is actually not the case.

- Assume no drop-offs take place within the 120 seconds of the trip ends.

Since drivers may spend more than 2 minutes in a large parking lot or garage in order to find a parking space, this type of parking activities at the end of a trip maybe detected as a trip-chaining stop using the criteria listed above. To avoid this, it is assumed no drop-offs take place within the 120 seconds of the trip ends.

Durations of engine-off and engine-on stops are calculated based on the number of GPS records captured during the stopping period.

Identify Different Routes Chosen by a Certain Driver

Due to the large size of data in this dissertation (1820 commute journeys), manually comparing the different routes (identified in the map-matching process) traveled by the same driver is time consuming. The author developed a script to compare all the commute routes of the same driver. A route is represented by a sequence of network links. If the network links of two routes share greater than 90 percent of the total length, they are assumed to be the same route; different otherwise. As drivers' behaviors do vary a lot, the route choice patterns are extremely complex. The author performed a manual double-check of the automatic route choice pattern detection result on the commute routes that were identified as different. Minor deviations around the neighborhood streets close to trip ends or deviations to avoid a certain intersection at a network node are not counted as route change.

Commute Duration and Travel Duration Calculation

In this dissertation, commute duration is defined as the total time elapsed between the time point when the driver turns on the vehicle engine and leaves home and the time point when he/she turns off the vehicle engine arriving at the work place. Travel duration equals the commute duration minus all the stop durations during that commute journey.

Travel Distance Calculation

Travel distance is calculated by accumulating the second-by-second linear distance between two consecutive GPS points based on the following function:

$$distance = \sqrt{(lat1 - lat2)^2 + (long1 - long2)^2} \tag{5.1}$$

The GPS data in decimal degrees format and position are stored in latitude and longitude format. Because one decimal degree corresponds to different distance values at different latitudes, directly summarizing the distance values in decimal degrees format for positions at different latitudes introduces error. This dissertation developed a script that converts the distance from decimal degrees into feet based on the conversion values in Table 5.1, and then calculates the distance using Equation 5.1. Because the GPS position accuracy is determined by the GPS data quality, only positions of valid GPS points (defined by number of satellites and PDOP values) are used in distance calculation. In case of invalid GPS point, linear distance is calculated based on the valid GPS point immediately before and after the current invalid one.

Table 5.1: Conversion Factor of Deferent Latitude Values

Latitude	0	5	10	15	20	25	30	35	40	45
Feet per second	101.45	101.07	99.92	98.02	95.37	92	87.93	83.2	77.83	71.86
Latitude	*50*	*55*	*60*	*65*	*70*	*75*	*80*	*85*	*90*	
Feet per second	65.34	58.32	50.85	42.99	34.8	26	17.68	8.87	0	

Commute Information Database

The serials of data processing algorithms generate a very rich set of aggregated data from the original GPS trips for further research analysis. This dissertation set up an Access database to organize all the commute journey level information and the driver and household information. The commute information database structure is shown in Figure 5.2.

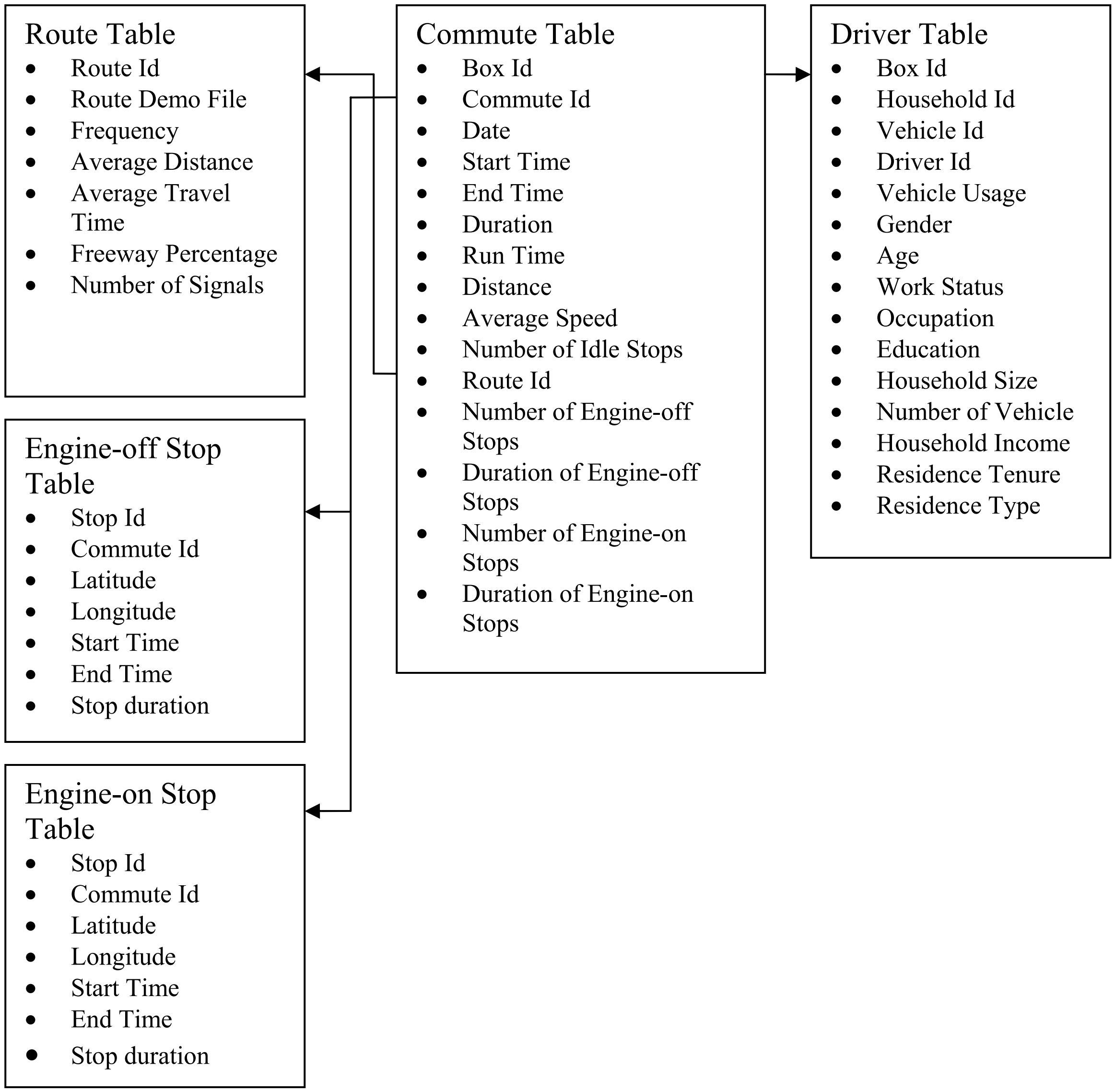


Figure 5.2: Commute Information Database

Chapter 6

Morning Commute Trend

The work journey is often the longest distance most people travel on a daily basis. Commute patterns are affected by workers' characteristics, location of home and work places, and the time and modes of commuting. Transportation practitioners seek means of inferring travel patterns from readily-observable variables, such as those available from the Census so that future travel behavior could be predicted with appropriate behavioral models using forecast of Census-type variables.

General Commute Trend in the Metro Areas of the United States

Source: Journey to Work Trends, [2004]

Population Change

According to Journey to Work Trends, the U.S. population grew at an unexpected pace between 1990 and 2000, adding 32.7 million people (13.2 percent) over the ten-year period. This represents the largest numerical increase in population in any decade in American history. Urbanization continued in the 1990s at the national level with over 80 percent of the population living in metropolitan areas. The suburban counties of major Metropolitan Statistical Areas (MSAs) incurred growth in area, population, and workers.

Household Composition

Average household size decreased from 3.3 persons per household in 1960 to 2.6 persons per household in 2000. Household composition also undergoes big changes. Married couples, with or without children, have become less common in the U.S. For the first time, the proportion of single-person households (25.8 percent) is greater than the number of nuclear families (married couples with children) which is 24.3 percent. The share of family households (a family household is composed of at least two people related by birth, marriage, or adoption) fell from 81 percent in 1970 to 68 percent in 2000. Non-family households (non-family households are a mix of people living alone, unmarried couples, and people live with friends or roommates) grew from 19 percent of the 1970 to 31.9 percent in 2000. Household composition is a major influence on household travel behavior. The change in households from a traditional nuclear family to more diverse and smaller households adds to the number of people who travel separately to work.

Worker Demographic

Change in worker demographic can have a strong impact on commute behavior.
The number of workers in the U.S. has doubled since 1960, from 65 million to 128 million. The large additions to the U.S. workforce seen every decade since 1960 may be near an end as the baby boomers enter into retirement, but immigration can be a factor that fills the workforce. Policy decisions determine the amount of immigration each year, if the trend continues, foreign-born people will be a large factor in population and worker growth in the U.S. One dramatic change in the workforce is the inclusion of women. In 1960, only 32.3 percent of the workforce is women, compared to the 46.7 percent in 2000.

Vehicle Ownership and Vehicle Use

Vehicle per household rose from 1.0 to 1.7 from year 1960 to 2000. In 2000, only around 10 percent of the households had no vehicle. With the decreasing household size, even fewer people are affected. On the other hand, around 55 percent of the households have two or more vehicles in year 2000 compared to 21 percent in 1960. In 1960, 43 million workers commuted by private vehicle, compared to 97 million workers commuting by private vehicle in the year 2000. The census shows that in year 2000, 75.7 percent of commuters drove alone to work, followed by carpooling (12.2 percent), transit (4.7 percent), work at home (3.3 percent) and walk (2.9 percent). Between 1990 and 2000, drove alone continued to increase, as carpools continued to drop. By year 2000, the average vehicle occupancy for the commute trip was 1.08.

Significant Increases in Commute Time

American workers are spending more time on their commute. In 2000, the average travel time to work was 25.3 minutes. While the average commute increased by 3.1 minutes between 1990 and 2000, there was only a 40 seconds increase from 1980 to 1990. In 2000, 14 percent of workers traveled more than 45 minutes compared to 12 percent in 1990, and 29 percent commute less than 15 minutes, compared to 31 percent in 1990. Forty percent of the commuters in large metro areas travel over 30 minutes to work, one way, on an average day.

Atlanta Metro Area Commutes

A metro area is consists of a core county / core counties containing a city of population greater than 50,000 people or a census defined urbanized area. Outlying counties are added to the metro area based on population density and commute behavior. Based on the 1999 Census definition, the Atlanta Metro Area contains 18 counties. The Atlanta metro area is among the ten fastest growing metropolitan areas during 1990 and 2000; the population increased 38.9 percent during the ten-year period.

One of the biggest changes in the worker flow patterns in Atlanta has been the huge increase in the number and percent of workers commuting between suburban residence and suburban work place. In year 2000, 53 percent of the commuters traveled suburban-to-suburban (compared to the 35 percent in 1970). Suburban-to-central is 20 percent, central-to-central is 13 percent (compared to 29 percent in 1970). Central-to-suburban is 5 percent, and other patterns account for 9 percent.

Travel time for commuters grew dramatically due to the immense population and worker growth. From 1990 to 2000, workers in Atlanta experienced the highest increase in travel time (5.2 minutes compared to the 3.1 minutes national average). From 1980 to 2000, the percent of workers with short commutes reduced dramatically in the suburban and ex-urban area, while the percent of workers with longer commutes increased dramatically in all three areas. Figure 6.1 shows the commute time distribution of Atlanta MSA in Census 1990 and Census 2000. The figure shows a decrease in the percentage of workers who had travel time less than 30 minutes, and an increase in the percentage of workers who had travel time longer than 45 minutes.

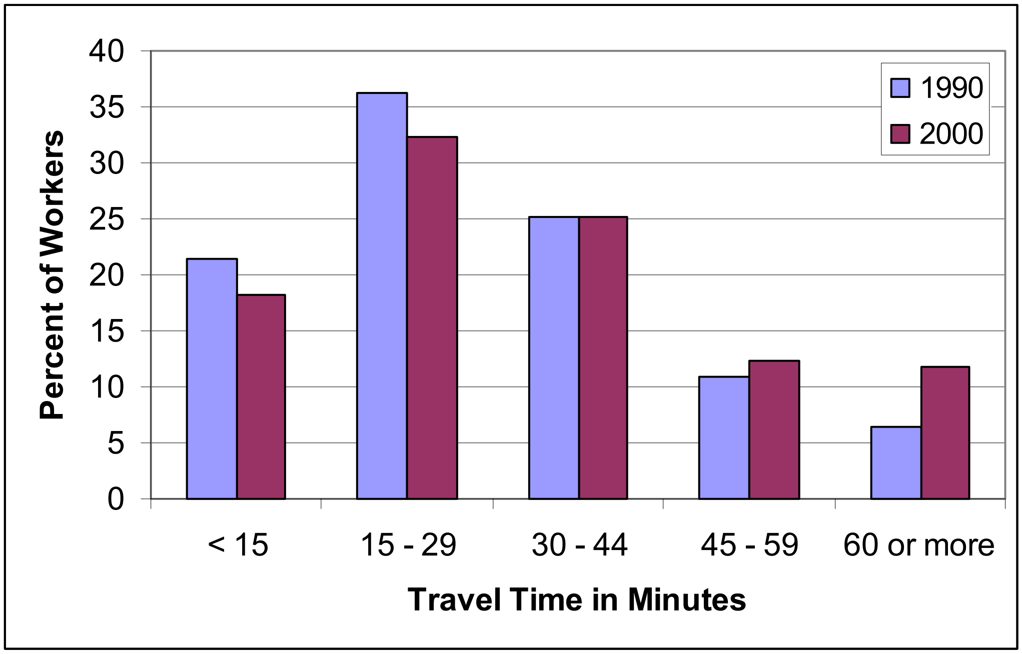


Figure 6.1: Atlanta MSA Commute Time Distribution

(Source: Journey to Work)

Atlanta is a fast-growing city with highway-oriented development. The private vehicle, especially driven alone to work, is the mode of choice for most Atlanta commuters. The percent of workers driving alone to work increased during the years. Mode split in year 2000 is shown in Figure 6.2. Around 3 out of 4 commuters drive alone to work, and transit is utilized by only 3 percent of the commuters.

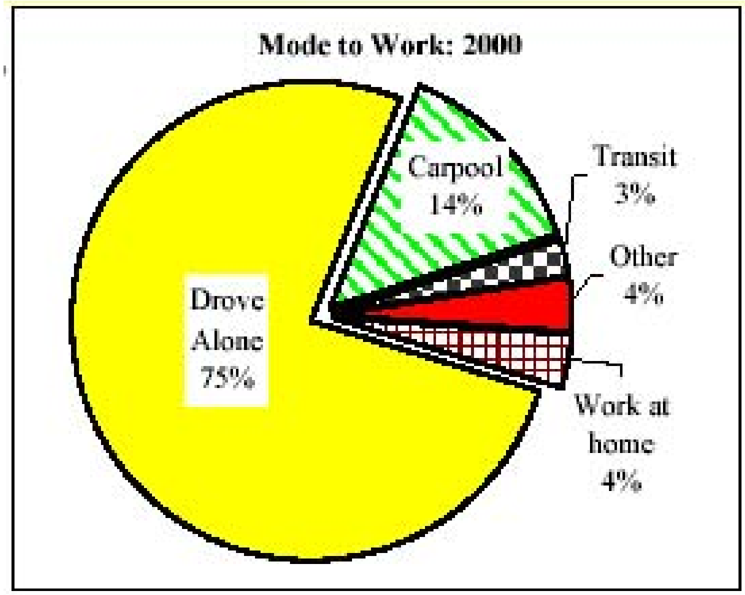


Figure 6.2: Commute Mode Split in Atlanta MSA

(Source: Journey to Work)

The highest percent of workers leave between 7:00 and 8:30 a.m. (see Figure 6.3). Comparing the distribution of year 1990 and year 2000, we can see evidence of peak spreading. There is a slight shift to earlier departures from 1990 to 2000.

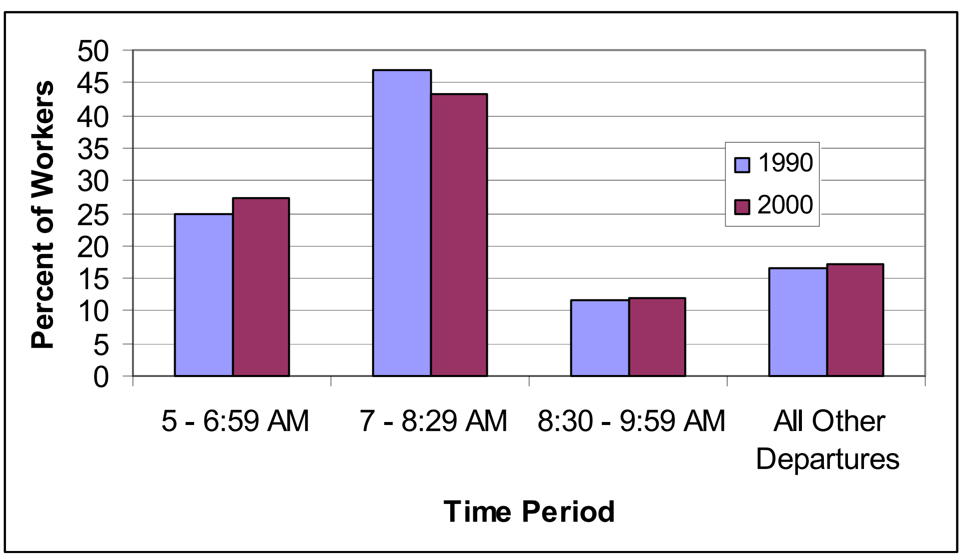


Figure 6.3 Atlanta MSA Morning Commute Departure Time

General Commute Pattern in the Sample

The majority of the Commute Atlanta data collection started in August, 2003. The data set used in this dissertation was created in December, 2003. The total duration of the data used in the sample is around 3 month to avoid the summer season and the holiday season. Specific observation period for each commuter may vary due to the fact that not all the data collection instruments were installed at the same time. In order to capture as many commute behavior during a representative time period as possible, the author decided to use ten-day worth of morning commute data as the study duration. Under these criteria, commute behaviors from 182 drivers were identified. If the study period is too long, fewer drivers meet the duration requirement. If the study period is too short, the data set may not be adequate to represent general commute behavior of that driver. Due to the fact that a certain driver may not necessarily work all five work days and the driver may use travel mode other than drive alone using his primary car occasionally, the 10-day period does not necessarily represent 2 work weeks (Monday through Friday).

To meet the research goal of this dissertation, only the 182 drivers that have known gender information, work full time or part time at a fixed working location, do not share their vehicle with another household member, and commute from home to work during the morning commute period are included in the data subset. Significantly fewer commuters in the lower income households meet all of these conditions. The household recruitment strata in the Commute Atlanta study are based on annual household income, household size and vehicle ownership. The household recruitment strata, and the subset of these households used in the analyses reported herein, are provided in Table 6.1.

The Commute Atlanta samples are slightly skewed to the higher income groups comparing to the Atlanta population due to the restrictions in vehicle ownership and vehicle sharing. This difference is expected since the objective of the Commute Atlanta project is designed to access effects of by-the-mile congestion pricing on commute travel behavior, and only households that own vehicles were recruited. The researchers also found out that higher-than-expected refusals and opt-outs of lower income households and higher-than-expected retention of upper income households. In the Commute Atlanta project, participants are not monetarily incentivized to participate, but the monitoring devices provide participants with vehicle theft tracking capabilities. Upper income households may have also placed a higher value on the Commute Atlanta project objectives (specifically on the identification of congestion locations). Plus, the trust placed in the researchers to maintain confidentiality of such revealing data may have differed across strata. Details on the recruitment process and study refusal rates are detailed in Ogle, et al. [2004].

The households in the dissertation sample are slightly skewed to the higher income groups comparing to the Commute Atlanta sample. It is possibly due to the fact that the commuters with white collar occupations usually have higher salary and fixed working schedule. On the other hand, the commuters with blue collar occupations who work at shifts may have commute schedules different from the traditional morning and afternoon peak time periods studied. Hence household income values for the commuters identified during the morning peak periods are higher than the overall working population. The net result, however, is that upper-income households and more educated individuals are over-represented in the sample when compared to census demographic profiles of the Atlanta MSA population. Hence, conclusions regarding behavior with demographics need to be restricted to each sample strata where sufficient data are available.

Table 6.1 Household Recruitment Strata

Sampling Group	*Annual Income*	*House-hold Size*	*Vehicle per HH*	*Atlanta Population Percent*	*Commute Atlanta HH Sample Target*	*Commute Atlanta HH* Recruited *(percent)*	*HH in the Dissertation (percent)*
0	Any	Any	0	7.4%	0	0 (0%)	0 (0%)
1	<$30,000	Any	1+	18.4%	35-40	20 (7.46%)	4 (2.90%)
2	$30,000 - $75,000	1	1+	11.3%	35-40	34 (12.69%)	17 (12.32%)
3	$30,000 - $75,000	2+	1	6.8%	35-40	18 (6.72%)	7 (5.07%)
4	$30,000 - $75,000	2	2+	10.6%	35-40	38 (14.18%)	13 (9.42%)
5	$30,000 - $75,000	3+	2+	13.9%	35-40	34 (12.69%)	14 (10.14%)
6	$75,000+	1	1+	2.8%	0	5 (1.87%)	4 (2.9%)
7	$75,000 - $100,000	2+	1+	12.1%	35-40	41 (15.30%)	26 (18.84%)
8	$100,000 +	2+	1+	16.8%	35-40	73 (27.24%)	51 (36.96%)
99	Unknown	Any	Any	na	0	5 (1.87%)	2 (1.45%)
Total				**100%**	**280**	**268 (100%)**	**138 (100%)**

Socio-demographic Characteristics

The summary of the sample socio-demographic statistics is shown in Table 6.2. Socio-economic characteristics have an important effect on commuter's travel behavior. A rich source of individual and household data has been collected in the household travel survey. The average household size is 2.86 persons per household. The average age of the drivers is 43 years. Most of the drivers have resided at their current residence location for more than 3 years, indicating a good level of familiarity with their travel areas. The respondents are divided fairly equally between males and females with 49.5 percent being males. Children less than 16 years of age are present in 52 households (70 commuters) and children 5 years or younger are present in 20 households (25 commuters). The ratio of workers per household is 1.45, which is comparable to 1.37 from the Census 2000 data for Atlanta MSA. Household vehicle ownership of the sample is higher than the average value of the Census 2000 for the Atlanta MSA (2.37 vehicles per household compared to the 1.8 vehicles per household of the Census 2000). At least 55 percent of the drivers in the sample have either undergraduate or postgraduate educations. The median household income of the sample is between $75,000 and $99,000. Household income in the sample is significantly higher than the median household income of the Atlanta MSA ($51,948 in the year 2000 Census).

Table 6.2: Sample Socio-demographic Characteristics Summary

Average household size	2.86
Commuters' residence type	
Single house	162 (89.01%)
Apartment, condo or townhouse	11 (6.04%)
Unknown	9 (4.95%)
Commuters' tenure at residence	
Less than one year	4 (2.20%)
One to three years	27 (14.84%)
More than three years	142 (78.02%)
Unknown	9 (4.95%)
Percent of male / Percent of female	49.45 / 50.55
Average number of vehicles in household	2.37
Number of households with children younger than 16	52
Number of households with children younger than 6	20
Average number of fulltime workers	1.45
Commuters with education of	
College graduate and above	100 (54.95%)
Not college graduate	63 (34.62%)
Unknown	19 (10.44%)
Commuters from household income group	
Less than $10,000	0 (0%)
$10,000 – 19,999	1 (0.55%)
$20,000 – 29,999	4 (2.20%)
$30,000 – 39,999	7 (3.85%)
$40,000 – 49,999	12 (6.59%)
$50,000 – 59,999	18 (9.89%)
$60,000 – 74,999	18 (9.89%)
$75,000 – 99,999	40 (21.98%)
$100,000 and above	78 (42.68%)
Unknown	4 (2.20%)
Commuters from age group	
(cut-off points are based on the census age groups)	14 (7.69%)
Under 25	33 (18.13%)
25 – 34	37 (20.33%)
35 – 44	54 (29.67%)
45 – 54	38 (20.88%)
55 – 64	3 (1.65%)
64 +	3 (1.65%)
Unknown	

General Commute Characteristics

- Travel Time

The mean travel time (exclude stop and drop-off time) of the sample is 31.53 minutes. Travel time distribution is shown in Figure 6.4. The null hypothesis that travel time of the sample has the same distribution compared to the Census 2000 in Figure 6.1 was not rejected at the 0.05 significance level by the chi-square test.

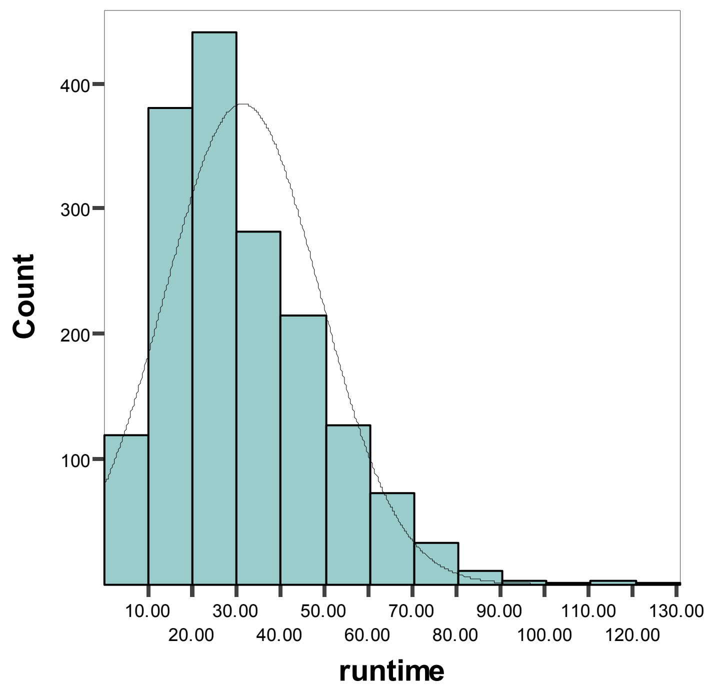


Figure 6.4: Morning Commute Travel Time Histogram

- Commute Duration

The mean commute duration (include stopping time) of the sample is 38.18 minutes. Commute duration distribution is shown in Figure 6.5. A few commutes with long durations mainly due to long stop duration on way of work skewed the distribution.

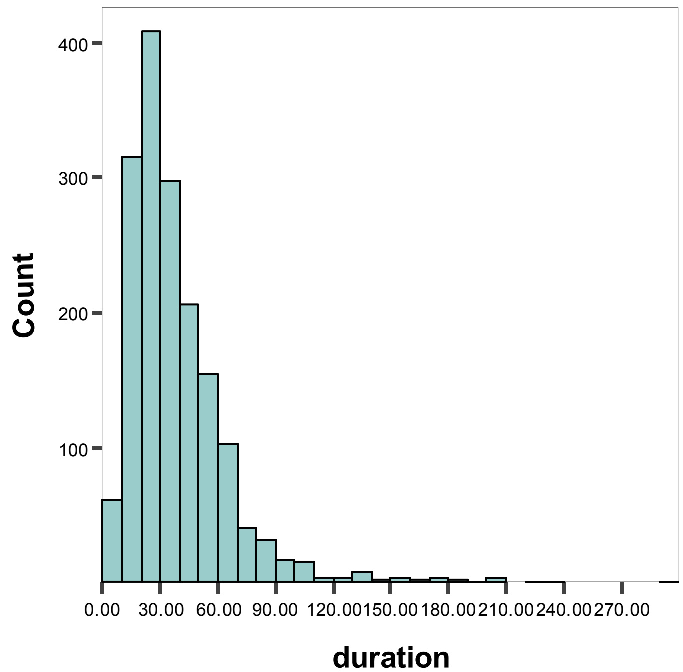


Figure 6.5: Morning Commute Duration Histogram

- Travel Distance

The mean travel distance is 16.35 miles. Travel distance distribution is shown in Figure 6.6. The distribution of the commute distances indicated that a large percent of the trips are short distance commute; i.e. more than 15 percent of the commutes are less than 5 miles, and around 80 percent of the commutes are less than 25 miles. Only 2 out of the 182 commuters have general commute distance longer than 50 miles.

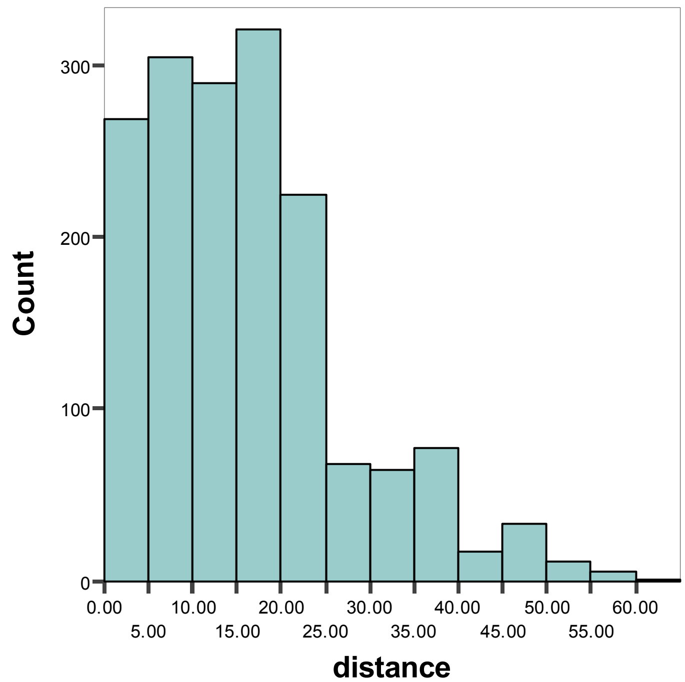


Figure 6.6: Morning Commute Travel Distance Histogram

- Departure Time

Morning commute departure time distribution is shown in Figure 6.7. The highest percent of workers leave between 7:00 a.m. and 7:59 a.m. The null hypothesis that departure time of the sample has the same distribution compared to the Census 2000 based on the Census category in Figure 6.3 was not rejected at the 0.05 significance level by chi-square test.

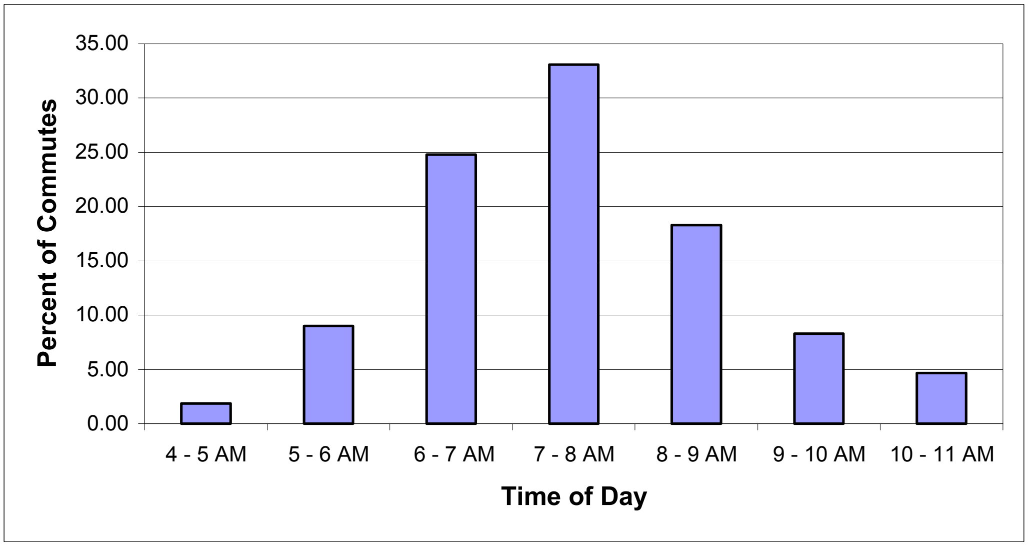


Figure 6.7: Morning Commute Departure Time Distribution

- Trip-Chaining

The work place anchors some of the non-work travel, either in intermediate stops commuters make between home and work or in trips around the workplace. As the literature pointed out, a secondary role of the commute journey is to provide an opportunity to link non-work travel with the commute itself [Nishii et al., 1989]. Commuting trips are becoming more and more complex as workers incorporate personal, household, and child-care activities into their commutes [Bianco and Lawson, 1996]. For example, Orski [1989] found that more than 60 percent of office workers who drive their personal car to work made intermediate stops on the way to or from work at least three times a week. Li et al. [2003] found out approximately 60 percent of a sample of 56 commuters stop on their way to work at least one day during their 5-day commute period and more than 15 percent of the drivers stop every day during their morning commute journey. Davidson [1991] also found that employees were twice as likely to make stops on their way home from work as on their way to work from home. The need to make stops on the way is one frequent reason for changing routes and was cited by 15.5 percent of respondents in the study by Abdel-Aty et al. [1994].

In this dissertation, trip-chaining stops made during the morning commute have been divided into two types: Engine-off stops take place when the driver turns off the engine during the stop. Engine-on stops take place when the driver does not turn off the engine during the stop (drop-off or pick-up). An example of the engine-off stop (at a day care center) and one engine-on stop (at a video store) is shown in Figure 6.8.

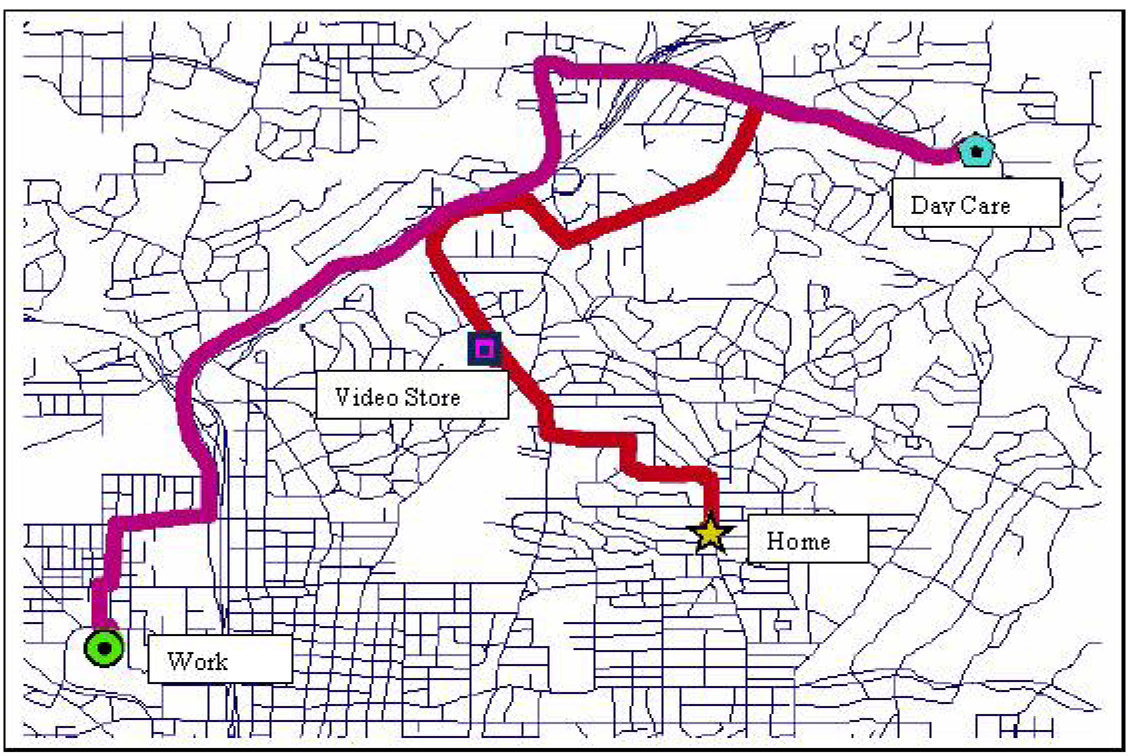


Figure 6.8: An Example of Morning Commute with Trip-Chaining

The frequency of non-work stops during the morning commute is shown in Table 6.3. 537 out of 1820 (30.5%) morning commute journeys have one or more stops. Similarly, Hanson [1980] found a 29.4 percent of passenger vehicle trips having one or more stops between home and work. In a survey of 164 respondents, Mahmassani et al. [1996] found 24.3 percent of the morning commute trips have one or more stops. Compare of these numbers shows that GPS-based data collection methods may be more effective in capturing trip-chaining behavior. Yalamanchili et al. [1999] compared the trip-chaining indications provided by the GPS data with those provided by the recall data. Results of their study showed that the GPS-based data performed in a superior manner to the recall data in capturing multi-stop chains in that the former captured more than twice as many multi-stop chains as the latter when comparisons were made in the context of a one-day travel period.

Table 6.3: Number of Stops on Morning Commutes

Number of stops	*Frequency Engine-off*	*Percent Engine-off*	*Frequency Engine-on*	*Percent Engine-on*	*Frequency Engine-off & Engine-on*	*Percent Engine-off & Engine-on*
0	1440	79.12	1593	87.53	1283	70.49
1	314	17.25	200	10.99	404	22.20
2	53	2.91	23	1.26	94	5.16
3	12	0.66	3	0.16	32	1.76
4+	1	0.05	1	0.05	7	0.37
Total	**1820**	**100**	**1820**	**100**	**1820**	**100**

Chapter 7

Morning Commute Route Choice Pattern

Some commuters utilize only a single route for their morning commute; others select routes from a choice set. For the numerous potential routes between an origin and destination pair, some routes share links, others have no overlap. GIS systems together with GPS data can reveal this important spatial pattern of route choice that was impossible to discern with earlier conventional survey methods such as interviews, respondent-administered questionnaires, and driver simulators. This chapter summarizes general findings of morning commute route choice patterns including the number of commute routes for each driver and the spatial deviation pattern of a commuter's different routes.

Number of Commute Routes

If one defines the most frequently used route between an O-D pair during the study period as a commuter's primary route, a total of 1528 out of 1820 (84 %) of the morning commute journeys were on the primary routes. The remaining 292 (16 %) commutes were on the alternative routes. In the sample, around 40 percent of the commuters had only one route for commutes during the 10-day period studied (see Table 7.1). The remaining 60 percent of the commuters used at least two routes for their commute. If we define the routes that appear at least twice during the study period as routine routes, around two thirds of the commuters have one routine route and one third of the commuters have two routine routes (see Table 7.1). Very few commuters have more than

2 routine routes. This result is higher than the research result of Abdel-Aty et al. [1994], in which only 15.5 percent of the respondents said they use more than one route to work. The difference may result from the fact that commuters may consider several similar routes as a single one in a travel survey.

Table 7.1: Number of Commute Routes Distribution

Number of Routes	*Number of Commuters*	*Number of Routine Routes*	*Number of Commuters*
1	72 (39.6%)	1	122 (67.0%)
2	55 (30.2%)	2	54 (29.7%)
3	35 (19.2%)	3	6 (3.3%)
4	16 (8.8%)	4	0 (0.0%)
5	4 (2.2%)	5	0 (0.0%)
Total	182 (100.0%)	**Total**	182 (100.0%)

Shared Route Distance

Shared route distance is defined as the distance sum of the route sections that are shared by all the commute routes of the same driver. The percentage of shared distance divided by the mean commute distance for the same driver is calculated as a measure of route deviation. The frequency of commuters fall in each shared distance category is shown in Figure 7.1. The extent of sharing ranges from 0 percent when all the commute routes of the same driver are completely different to 100 percent when only one route exists for the driver. The figure shows that most commute routes of the same driver share at least some network links.

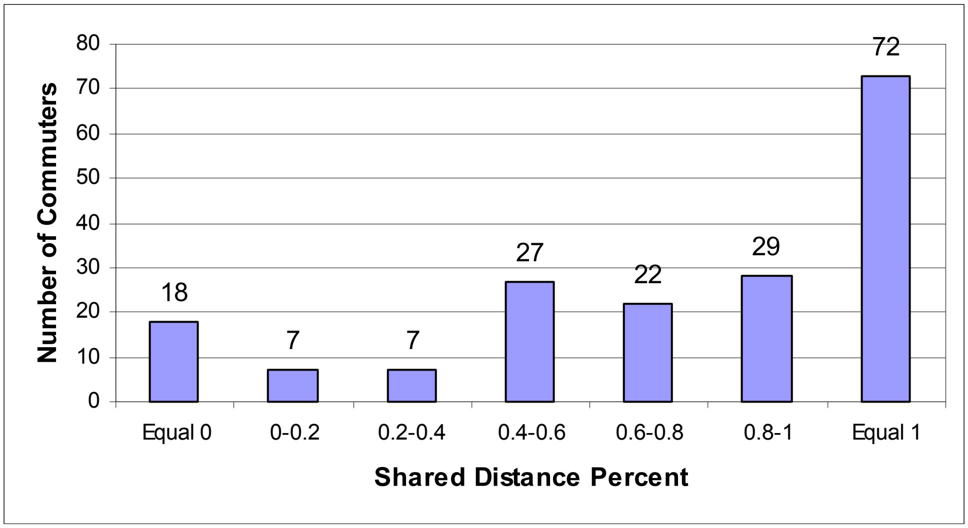
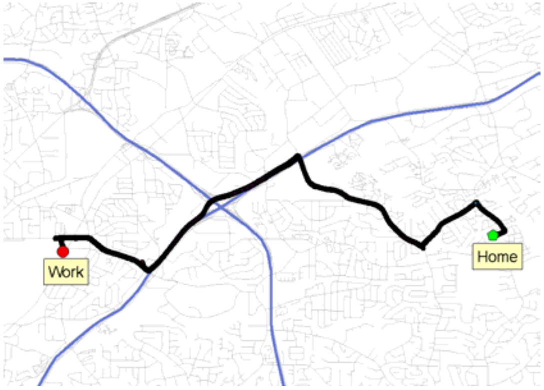


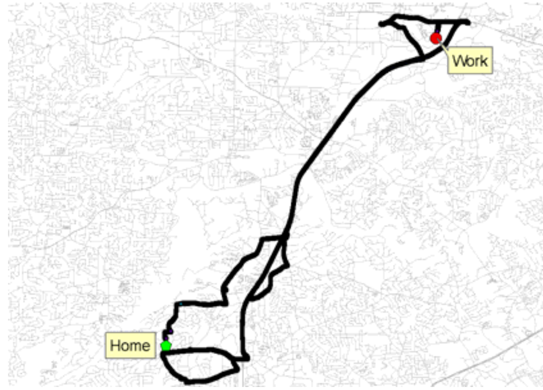
Figure 7.1: Shared Distance Distribution

Route Deviation Patterns

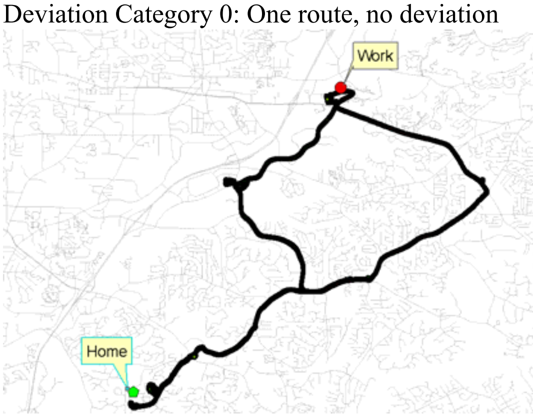
Depending on the driver's familiarity of the road network, deviation can occur anywhere along the route. One shortcoming of the percentage of shared link length is that it cannot reveal the spatial deviation pattern. Hence, this dissertation defines a group of eight deviation patterns, shown in Table 7.2, to differentiate those pairs of routes that do not share all the links. Among the 182 commuters, 72 of them used only one commute route during the 10-day study period. Eighteen commuters used routes that are completely different. The remaining commuters' routes deviate close to either home or work, or in the middle of the route. Visual examples of each category are shown in Figure 7.2.



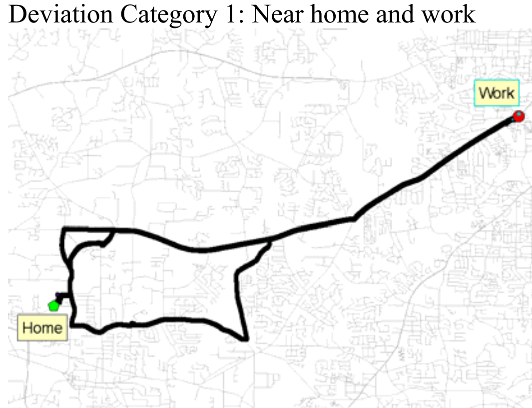
Deviation Category 0: One route, no deviation



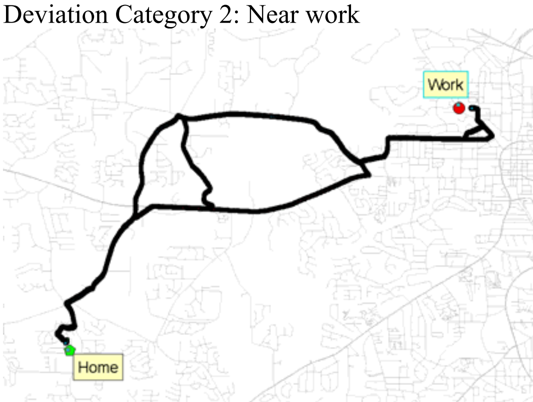
Deviation Category 1: Near home and work



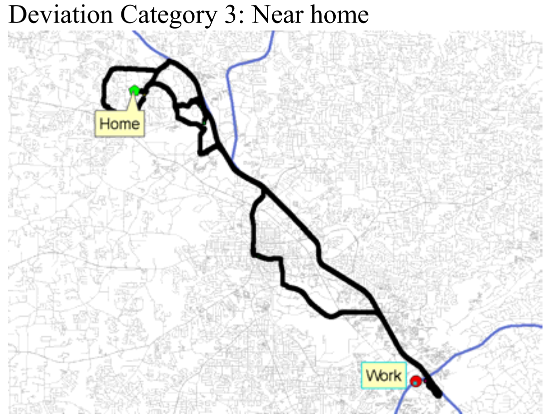
Deviation Category 2: Near work



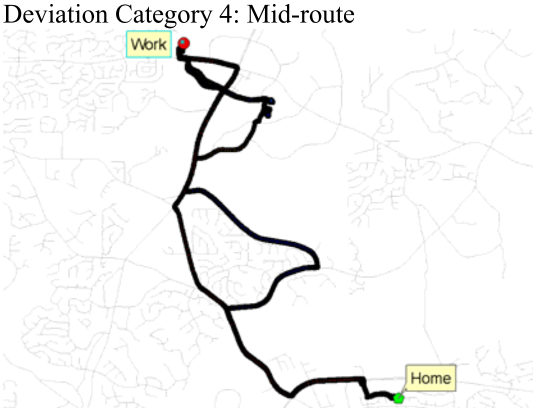
Deviation Category 3: Near home



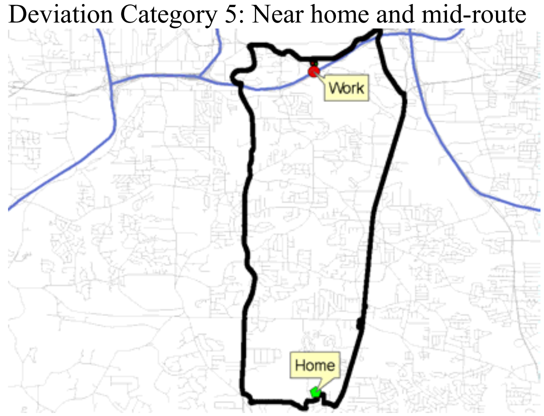
Deviation Category 4: Mid-route



Deviation Category 5: Near home and mid-route



Deviation Category 6: Near Work and mid-route



Deviation Category 7: Complete different

Figure 7.3: Visual Examples of Route Deviation Patterns

Figure 7.3 shows the average shared distance percentage by route deviation pattern. Except the routes that are completely different, deviations along the middle of the routes have lower shared distance percentage comparing to deviations close to trip origins and destinations.

The null hypothesis that the shared distance percentages are equal for all the deviation patterns (except pattern 0 and 7) is rejected by the ANOVA test at 0.05 significance level, which also indicates that a significant relationship between commute distance and shared link percentage does exist. The Tukey Kramer multiple comparison test shows that groups 4,5,6 (deviations taking place in the middle of the routes) are significantly different from groups 1,2,3 (deviation taking place near the home or work ends) at 0.05 level.

Table 7.2: Route Deviation Categories

Deviation Code	*Deviation pattern*	*Frequency*
0	No deviation, one route	72
1	Deviation near home and work	10
2	Deviation near work place	21
3	Deviation near home	20
4	Deviation mid-route	30
5	Deviation near home and mid-route	5
6	Deviation near work and mid-route	6
7	Complete different	18

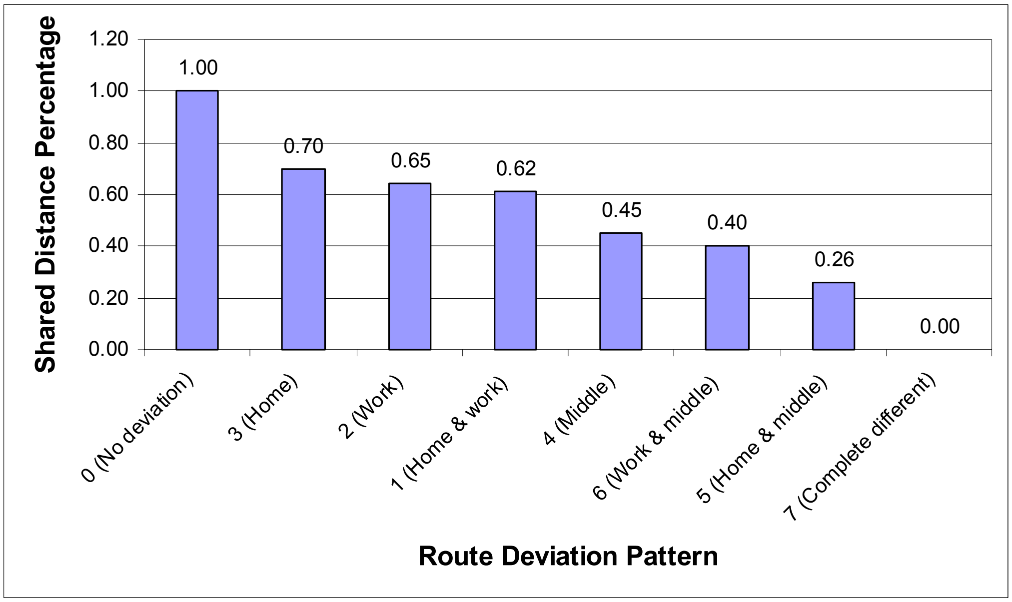


Figure 7.3: Shared Distance Percentage by Route Deviation Pattern

Table 7.3: ANOVA Test of Shared Distance Percentage of Different Deviation Patterns

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Shared distance percent	Between Groups	1.224	5	0.245	4.934	.001
	Within Groups	4.216	85	.050		
	Total	5.440	90			

Commute Distance and Deviation Pattern

Distance is expected to play an important role in commuters' route choice. As distance increase, more reasonable paths exist that connect the origin and the destination. Hence, the traveler has more opportunity to deviate. On the other hand, as distance increases, driver's familiarity with the network decreases. A driver normally knows well the areas close to the home and the work locations, but may not be familiar with the road network in the middle of the route. The relationship between deviation pattern and commute distance is shown in Figure 7.4. The figure shows, when commute distance is relatively short, drivers either choose a single route, or choose completely different routes. As distance increases, deviations occur all along the route, either mid-route or close to trip ends. When commute distance is relatively long, most of the deviation occurs close to the home end.

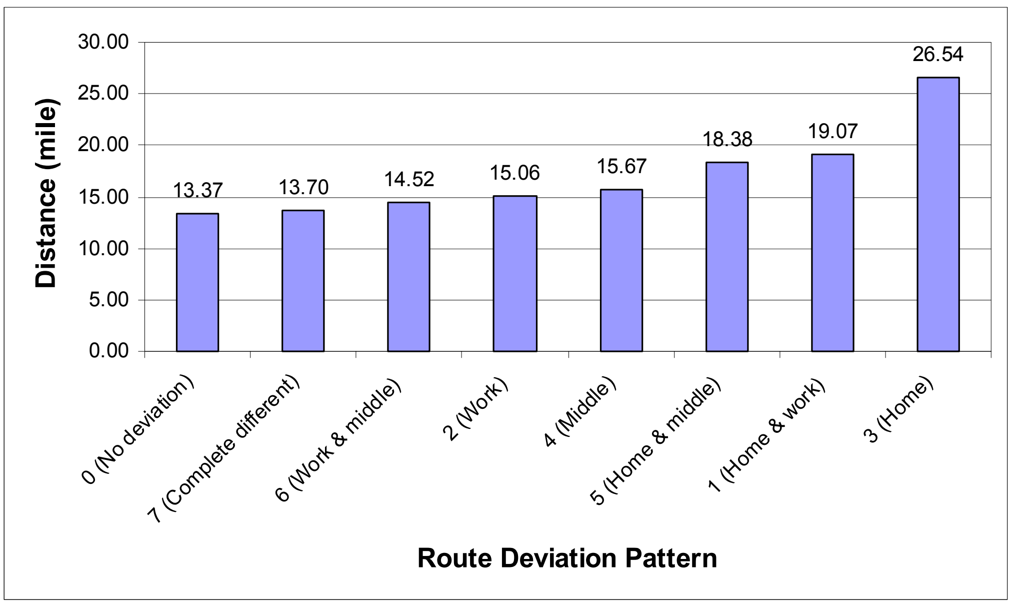


Figure 7.4: Deviation Pattern and Mean Commute Distance

The null hypothesis that the mean commute distances are equal for all the deviation patterns is rejected by the ANOVA test at 0.05 significance level, which indicates that a significant relationship between commute distance and route deviation pattern does exist. The Tukey Kramer multiple comparison test shows that group 3 (deviation near home) are significantly different from the other groups at 0.05 level (Table 7.4).

Table 7.4: ANOVA Test of Mean Commute Distance of Different Deviation Patterns

		Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Mean commute distance	Between Groups	2839.853	7	405.693	3.836	.001
	Within Groups	18295.263	173	105.753		
	Total	21135.116	180			

Chapter 8

Objective Route Choice Impact Factors

Objective route-level factors impact route choice decisions. Factors identified in Table 2.1, including road condition, traffic situation, and environment attributes, are the main research interest of this chapter. This chapter identifies the factors that appear to impact commuters' choice of primary routes identified in the study as their primary routes instead of all the other alternative routes identified. This chapter develops discrete choice models regarding the commuters' primary route and alternative routes as chosen and un-chosen routes, respectively. One hundred and ten commuters who used at least two different routes during the ten-day study period form the sample set for this analysis (see page 87 for a discussion of sample size determination).

Road Characteristics Information

Objective road characteristics used in the dissertation are based on the Georgia Department of Transportation (GDOT) Road Characteristics (RC) database for year 2000. This database contains road features collected by subcontractors of GDOT. The original database is in ACCESS format and each record is identified by a unique key composed of a RC link number and a mile-point number. In order to generate the road network of the study area that includes the RC information, the author combined the RC database with the GDOT Digital Linear Graph (DLG) GIS road network using linear referencing method. The method is based on the route system identifiable with the RC link number and the mile-point value in the DLG road network.

This dissertation incorporates the road characteristics that possibly impact commuters' route choice behavior, including road functional classifications and traffic controls. Road functional classifications in the RC database are shown in Table 8.1. Among them, interstate principal arterials, urban freeways, and expressways are evaluated as freeways. Traffic control devices are divided into the categories shown in Table 8.2. Among them, categories S, P and L are traffic signals.

No clear-cut accuracy measures of the RC database are available because the database is undergoing continuous updates and no accuracy information at a specific time is maintained. The author performed random accuracy check of the functional classification and traffic control device of the RC data in areas of Fulton, Cobb, Dekalb and Gwinnett counties. Based on the random check results, road functional classifications and speed limit information is mostly accurate. Traffic signal information is also mostly accurate. On the other hand, stop sign information is missing for some intersection locations. Among the 20 signals in the three locations, 18 (90%) of them were accurately represented in the RC database, and the remaining 2 (10%) were missing. The author also compared the signal information in the RC database with the traffic signal map for the City of Atlanta (See Appendix A). Traffic signals in the City of Atlanta were verified with the signal control map from the traffic & transportation office of the City of Atlanta. Among the 878 signals in the City of Atlanta, 716 (87.53%) of them were accurately represented in the RC database. The remaining 162 (12.47%) were missing in the RC database. Due to the low accuracy of the stop sign information, the author did not include stop sign in the further model development.

Table 8.1: Road Functional Classification Coding in the RC Database

Rural Categories	*Urban Categories*
Interstate principal arterial	Interstate principal arterial
Principal arterial	Urban freeway and expressway
Minor arterial	Urban principal arterial
Major arterial	Minor arterial street
Minor collector	Collector street
Local	Local

Table 8.2: Traffic Control Devices Coding in the RC Database

Signal Type	*Definition*	*Number of Signal in Database*
S	Traffic control device	2549
P	Traffic control with pedestrian signalization	1686
A	Stop sign	17909
F	Flasher, other than overhead beacon	299
L	Traffic control device with left turn arrow	951
B	Beacon, overhead flashing amber	181
R	Beacon, overhead flashing red	358
C	Stop, all directions	3788
Y	Yield sign	980
W	Yield sign, opposite direction of inventory	534
O	Stop sign, opposite direction of inventory	5319

Paired Sample T-Test

Paired sample t-tests that compare the differences in average route characteristics values between the primary route and the alternative routes of each commuter are shown in Table 8.3. Route characteristics analyzed include travel speed, travel distance, travel time, and road attributes such as number of signals and roadway functional classifications. Based on the GPS vehicle activity data, the number of idle stops is defined as the number of periods during the commute journey when the vehicle is traveling on the road network at a speed less than 5 mph for at least one minute duration. The variable is designed to catch the traffic volatility and driving experience on a certain route. A route with more traffic signals, stop signs and other traffic control devices, or a route is more congested may have larger number of idle stops.

All the one-tail t-statistics (because of assumed direction of change) are significant at the 95 percent confidence level. The results indicate that a certain commuter's primary routes have shorter running time, shorter distance, faster average running speed, fewer idle stops, fewer traffic signals, and higher percentage of freeway travel distance comparing to the alternative routes of that same commuter.

As shown in Table 8.3, the null hypothesis that the alternative routes have the same percentage of freeway distance is rejected by the paired sample t-test, which indicates that alternative routes employ a lower freeway percentage than primary routes. The result is comparable to the study result of Abdel-Aty et al. [1994]. In that study, secondary routes tend to have more surface streets than primary routes, possibly as alternatives used to avoid congestion on freeways.

Table 8.3: Paired Sample T-Tests of Primary and Alternative Route Attributes

	Mean Primary Route	*Mean Alternative Route*	*Mean Difference (Primary - Alternative)*	*t*	*df*	*Significance (2-tail)/(1-tail)*
Average Commute Duration (minutes)	38.81	54.34	-15.53	-4.700	109	0.000/0.000
Average Running Time (minutes)	32.80	39.10	-6.30	-5.427	109	0.000/0.000
Average Distance (mile)	17.12	18.96	-1.84	-4.116	109	0.000/0.000
Average Running Speed (mph)	29.25	27.90	1.35	2.214	109	0.029/0.014
Average Number of Idle Stops	2.43	3.27	-0.83	-4.715	109	0.000/0.000
Average Number of Signals	7.16	9.14	-1.98	-3.559	100	0.000/0.000
Average Percent of Freeways (between 0 and 1)	0.29	0.25	0.04	2.344	99	0.021/0.011

Table 8.4: Paired Sample T-Tests of Primary and Alternative Route Trip-Chaining

	Mean Primary Route	*Mean Alternative Route*	*Mean Difference (Primary - Alternative)*	*t*	*df*	*Significance (2-tail)/(1-tail)*
Average Number of Engine-off Stops	0.28	0.56	-0.29	-4.37	109	0.000/0.000
Average Number of Engine-on Stops	0.16	0.26	-0.09	-2.34	109	0.044/0.022

The average morning commute route contains 28.02 percent of distance traveling on freeways. Even in Atlanta, a city that is generally considered saturated with freeways, around 48 percent of the primary routes involve no freeway at all. Figure 8.1 summarizes percentage of freeway travel based on commute distance categories. The figure shows, as the commuter distance increases, the utilization of freeway also increases. For commute journeys shorter than 5 miles, only an average of around 2 percent of the distance was traveled on freeways. On the other hand, when the commute distance is longer than 25 miles, around 60 percent of the distance was traveled on freeway segments. The figure also shows that commute journeys with trip-chaining always have lower freeway percentage compare to commute journeys without trip-chaining of the same distance category. This result is not surprising since local streets provide stop opportunities on the way to work.

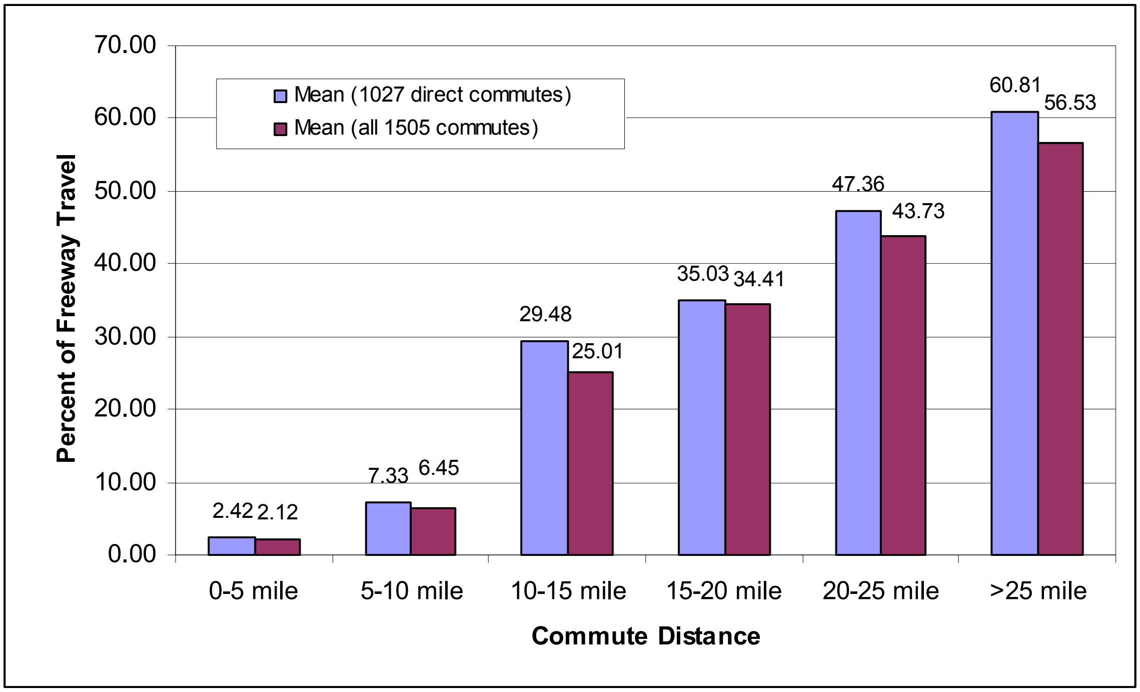


Figure 8.1: Commute Distance & Freeway Usage

A total of 537 out of 1820 (30.5%) morning commute journeys in the sample include one or more trip-chaining stops. The null hypothesis that the commuters who traveled on their primary routes have the same tendency to make an engine-off or engine-on trip-chaining stop comparing to the commuters who traveled on their alternative routes, is rejected by paired sample t-tests for difference in two means at the 0.05 significance level (see Table8.4). The result shows a correlation exist between trip-chaining and route choice behavior such that commuters who travel on their alternative routes have a higher likelihood of stopping on their way to work, or vice versa.

Conditional Logit Model of Objective Route Choice Factors

The decision of route choice is qualitative (or discrete) in nature. A route is chosen to the exclusion of one or more possible alternatives. This section presents Conditional Logit Model with multiple observations of the route choice based on the objectively measured route attributes.

Methodology Introduction

- Utility Theory and Multinomial Logit

For a discrete choice among J alternatives, assume U_{ij} represents the utility of the j^{th} choice to the i^{th} individual and U_{ij} can be represented by two components: $U_{ij} = V_{ij} + \varepsilon_{ij}$. The measurable part V_{ij} is a function of the measured attributes X. The other part ε_{ij} reflects the random part of the utility. Sources of this uncertainty come from unobserved alternative attributes, unobserved individual characteristics (unobserved personal preferences), measurement errors, and the use of instrumental variables [Simon et al., 2003]. If we assume that individuals act in a rational way and maximize their utility, subject i will choose alternative j if U_{ij} is the largest of $U_{i1}, \ldots, U_{iJ}$. The probability of individual i choosing alternative j from this set of alternatives (P_{ij}) is:

$$P_{ij} = P(U_{ij} >= U_{it}), \ \forall\, t \in J \text{ or}$$

$$P_{ij} = P(V_{ij} + \varepsilon_{ij} >= V_{it} + \varepsilon_{it}), \ \forall\, t \in J$$

If the representative utility of each choice depends only on the attributes of the decision maker, a measurable part of the utility $V_{ij} = X_i\beta_j$. If assuming random errors are independent and identically distributed (i.i.d.) Gumbel, the multinomial logit probability of individual i choosing alternative j from this set of alternatives (P_{ij}) is:

$$P_{ij} = \frac{\exp(\beta_j X_i)}{\sum_{t \in J} \exp(\beta_t X_i)}$$

In which, P_{ij} is the probability of individual i chose alternative j

$X_i\beta_j$ is the systematic portion of utility of alternative j for individual i

β is the estimation parameter

ε_{ij} is the random portion of utility of alternative j for individual i

- Conditional Logit Model

Conditional logit model is similar to ordinary logit model except that the data occur in groups. In the conditional logit model, an individual is faced with an array of choices and must choose one. It is an extension of the multinomial logit model in which the representative utility of each choice depends on choice specific attributes. Alternative-specific variables that vary by outcome and individual are used to predict the outcome that is chosen [Long, 2001]. The name conditional logistic regression for matched case control groups is often used by Biostatisticians and epidemiologists. Economists and other social scientists call this type of model as fixed-effects logit for panel data. It is also called McFadden's choice model since McFadden [1974] proposed the conditional logit model that models the expected utilities in terms of characteristics of the alternatives rather than attributes of the individuals. The utility function is: $U_{ij} = \alpha Z_{ij} + \varepsilon_{ij}$. If one assumes random errors are independent and identically distributed (i.i.d.) Gumbel, the conditional logit probability is:

$$P_{ij} = \frac{\exp(\alpha\ Z_{ij})}{\sum_{t \in J} \exp(\alpha\ Z_{it})}$$

In which, P_{ij} is the probability of individual i chose alternative j

αZ_{ij} is the systematic portion of utility of alternative j for individual i

α is the estimation parameter

ε_{ij} is the random portion of utility of alternative j for individual i

- Multinomial / Conditional Logit

In models of choice behavior, where the explanatory variables may include attributes of the choice alternatives as well as characteristics of the individuals making the choices. The predicted probability can be written as:

$$P_{ij} = \frac{\exp(\alpha\ Z_{ij} + \beta_j X_i)}{\sum_{t \in J} \exp(\alpha\ Z_{it} + \beta_j X_i)}$$

In which, P_{ij} is the probability of individual i chose alternative j

αZ_{ij} is the choice specific portion of utility of alternative j for individual i

$X_i \beta_j$ is the individual specific portion of utility of alternative j for individual i

α and β are the estimation parameters

ε_{ij} is the random portion of utility of alternative j for individual i

- Conditional Logit Model with Multiple Observations

The estimation of a discrete choice model based on multiple observations from the same object causes an obvious correlation of the disturbance term. In the multiple observation data, pseudo objects that share invariant unobserved factors make the potential problem of heterogeneity more severe. Heterogeneity is caused when behavioral differences are largely due to unobserved factors that are correlated with the observed explanatory variables. If heterogeneity is not taken into account, the model estimates will be biased and unstable. In previous research that used multiple observations from each object, Bunch et al. [1993] ignored the effect of heterogeneity by indicating that in a small number of multiple observations from each object the properties of parameter estimates do not rely on strict independence assumption. Louviere and Woodworth [1983] corrected the standard errors produced by a multiple responses regression model by multiplying the standard errors by the square root of the number of observations per object. The authors determined this method is conservative and overcorrect the value of the standard errors.

To obtain unbiased estimates of the explanatory variable, the effect of the correlation of disturbances must be taken into account when modeling discrete choice models based on multiple observations. Because the observation period of this study is within a relatively short time period (10 continuous days with observed commute activity), it is possible to argue that the values of most exogenous variable remain constant over time. Hence, the effects of those omitted variables that are specific to individual cross-sectional units stay constant over time. If α_i is added to the general utility function to account for the individual specific effects for individual i, the utility function is:

$$U^*_{ijt} = \beta X_{ijt} + \alpha_i + \varepsilon_{ijt}$$

In which, U^*_{ijt} is the utility of alternative j for individual i at time t

βX_{ijt} is the systematic utility of alternative j for individual i at time t

α_i is the individual specific effects for individual i

ε_{ijt} is the random utility of alternative j for individual i at time t

In this case, it is assumed that there are differences among the cross-sectional units but these differences stay constant over time and are not captured by the included explanatory variables X. When the individual-specific effects α_i are treated as fixed, the model is called fixed effects model. When α_i are treated as random, the model is called random effects model.

Model Estimation

A conditional logit choice model (or fixed-effect logit model for panel data) was set up using the methodology presented above. A commuter's primary route and alternative routes are defined as chosen and un-chosen route, respectively. To study the impact of trip-chaining on commuters' route choice decision, two sets of models were developed. Model group one is based on all the commute journeys from the drivers who used multiple routes during the study period and trip-chaining is modeled as an independent impact variable. Model group two is based on a subset of the data used in model one; only commute journeys with no engine-off or engine-on trip-chaining stops are included in the models.

Univariable conditional logit models were first fit for all the independent variables. The results are summarized in Table 8.5. In both models, all the independent variables have the expected sign and are statistically significant at the 95 percent confidence level. The pseudo R-square values indicate that travel time has the strongest explaining power, followed by travel distance.

Table 8.5: Univariable Conditional Logit Model Estimations

Variable	*Model group 1: All Commute Journeys*				*Model group 2: Direct Commute Journeys*			
	Coef.	*Odds*	*t (p)*	*Pseudo R^2*	*Coef.*	*Odds*	*t (p)*	*Pseudo R^2*
Travel Time	-0.0719	0.9306	-7.48 (0.000)	0.0873	-0.0702	0.9323	-3.86 (0.000)	0.0468
Distance	-0.2754	0.7593	-6.71 (0.000)	0.0803	-0.3024	0.7390	-3.15 (0.002)	0.0432
Speed	0.0534	1.0549	3.71 (0.000)	0.0170	0.0402	1.0410	1.80 (0.072)	0.0092
Number of idle stops	-0.3413	0.7108	-6.67 (0.000)	0.0594	-0.3506	0.7043	-3.59 (0.000)	0.0406
Percent of freeway	0.0391	1.0398	4.95 (0.000)	0.0356	0.0391	1.0399	3.20 (0.001)	0.0378
Number of traffic signals	-0.1556	0.8559	-6.23 (0.000)	0.0576	-0.1797	0.8355	-3.97 (0.000)	0.0593
Number of trip-chaining stops	-0.6521	0.5210	-5.72 (0.000)	0.0450				

In both model group one and two, travel time and travel distance are highly correlated [Table 8.6 & 8.7]. This multi-colinearity makes it impossible to evaluate the relative importance of each predictor. Hence, separate models were built based on either travel time or travel distance.

Table 8.6: Correlation Table of the Independent Variables (model 1 & 2)

	distance	*Travel time*	*Average speed*	*Number of idle stops*	*Percent of freeway*	*Number of traffic signals*	*Number of trip-chaining stops*
Distance	1						
Travel time	0.8179	1					
Average speed	0.6019	0.1317	1				
Number of idle stops	0.3794	0.7477	-0.3132	1			
Percent of freeway	0.5624	0.2867	0.6994	-0.0466	1		
Number of traffic signals	0.2995	0.3555	0.0557	0.3517	0.1003	1	
Number of trip-chaining stops	0.3202	0.4305	-0.0713	0.4713	0.0812	0.2076	1

Table 8.7: Correlation Table of the Independent Variables (model 3 & 4)

	distance	*Travel time*	*Average speed*	*Number of idle stops*	*Percent of freeway*	*Number of traffic signals*
Distance	1					
Travel time	0.8002	1				
Average speed	0.6738	0.198	1			
Number of idle stops	0.2976	0.6916	-0.2922	1		
Percent of freeway	0.5893	0.287	0.7355	-0.0613	1	
Number of traffic signals	0.3156	0.3579	0.1072	0.3311	0.1203	1

The final model specifications and parameter estimation results are presented in Table 8.8. The binary choice outcome in this study is always the primary route observed in our sample. For the alternative-specific variables, the odds ratios are the multiplicative effect of a unit change in a given independent variable on the odds of any given outcome. For example, based on case 1 of Table 8.8, increasing travel time by one minute for a given route decreases the odds of that route being chosen as the primary route by a factor of 0.9538 (4.62%), holding the values for the other alternatives constant.

Table 8.8: Conditional Logit Model Estimations

Variable	*Model Group 1 All Commute Journeys*		*Model Group 1 All Commute Journeys*		*Model Group 2 Direct Commute Journeys*		*Model Group 2 Direct Commute Journeys*	
	Coef. (Odds)	*t (p)*	*Coef. (Odds)*	*t (p)*	*Coef (Odds)*	*t (p)*	*Coef (Odds)*	*t (p)*
Travel time	-0.0473 (0.9538)	-4.43 (0.000)			-0.0421 (0.8588)	-2.01 (0.044)		
Distance			-0.2004 (0.8259)	-4.24 (0.000)			-0.3248 (0.7227)	-2.82 (0.005)
Number of idle stops			-0.1616 (0.8125)	-2.58 (0.010)			-0.1714 (0.8425)	-1.56 (0.118)
Percent of freeway	0.0182 (1.0184)	2.14 (0.033)	0.0239 (1.0262)	2.69 (0.007)	0.0117 (1.0118)	0.85 (0.397)	0.0146 (1.0147)	1.05 (0.292)
Number of traffic signals	-0.0796 (0.9235)	-2.92 (0.003)	-0.0594 (0.9433)	-2.11 (0.035)	-0.1236 (0.8838)	-2.40 (0.016)	-0.1055 (0.8998)	-2.12 (0.034)
Number of trip-chaining stops	-0.2896 (0.7486)	-2.16 (0.031)	-0.1012 (0.8962)	-0.70 (0.485)				
Model Summary Statistics								
Log Likelihood at Zero	-368.73		-368.72		-154.06		-153.76	
Log likelihood at convergence	-330.47		-326.34		-142.25		-136.96	
Prob>Chi2	0.0000		0.0000		0.0000		0.0000	
Pseudo R-square	0.1254		0.1363		0.0808		0.1150	
Number of observations	948		948		408		408	

Comparing the model results for model group 1 which used all the commute trips as the sample set with the univariate model results, we found out that travel time or travel distance alone can explain around 8 percent of the variation in the dataset. The models that include objective-level factors such as freeway percentage and number of traffic signals can explain around 13 percent of the variation in the dataset. Comparing the model results of model group 2 which used only the direct commute trips as the sample set with the univariate model results, we found out that travel time or travel distance alone can explain around 4 percent of the variation in the dataset. The models that include objective-level factors such as freeway percentage and number of traffic signals can explain around 10 percent of the variation in the dataset. Hence, including this type of information in the general route cost functions can help improve the predictability of the route choice models.

Disaggregate models using the individual level data inherently have more variability in the observed choices than the aggregated models using the averaged behavior data, so we do expect the pseudo R-squares to be smaller than the R-square values in the aggregated models. The low values of pseudo R-square in Table 8.8 may also possibly indicate that objective route attributes as a whole do not provide high explaining power for driver's route choice behavior. Among the above models, the model with the highest pseudo R square can only explain around 14 percent of the variability in the sample set. The unexplained part may be attributed to factors such as differences in perception, imperfect information on route costs. First, subjective perceptions and attitudes about objective route attributes on different alternatives drive the traveler to a certain choice. Different usage of route attributes, different perceptions of route attributes, different interpretation of the traffic network situation can result in different behavior even in the same situation. Second, travelers' perception of the relevant alternatives and their attributes is somewhat incomplete and inaccurate [Abdel Aty., 1995]. To some degree, the drivers' knowledge of attribute values is a distorted image of the actual value. Studies that collect both the objective level route choice factors through the field observation methods and the subjective level factors through stated preference surveys can further testify whether this is the actual case.

This chapter studies the route choice impact factors based on objective real-world observations of travel behavior during a multi-day period. The findings confirm that minimizing travel time, although very important, is not the only factor impact route choice. Several other factors have been identified to impact commuters' route choice. Those factors include traveling speed, driving experience in aspect of number of idle stops and traffic signals on the way, road functional classification in aspect of percentage of freeway travel distance. Including these factors in determining drivers' route choice behavior, and giving each factor a weight that represents its significance in the route choice will give us more realistic result.

Chapter 9

Choice of Single or Multiple Commute Routes

To assess commute route choice dynamics, this chapter models commuters' route choice decision during the study period using a single route or multiple routes. The dependent variable has binary outcomes that indicate whether a commuter used a single route during the 10-day period or used multiple routes. The model developed in this chapter is a behavioral model that is probabilistic, and based on the evidence of drivers' valuations of a number of route and trip characteristics as well as the commuters' socio-demographic characteristics.

Methodological Approach

The linear probability model assumes that probability P is a linear function of X.

$P = F(X, \beta) = X\beta$, where β is a column of parameters, and X is a matrix of observations on the explanatory variables. Although the linear probability model is computationally simple and has very little structure or assumptions imposed on the data, it has certain problems. One problem is, even though the mean of the error term of a linear probability model is zero, the variance of the error term is heteroschedastic. Observations for which $P_i = X_i\beta$ is close to 0 or 1 have relatively low variance while observations with $P_i=X_i\beta$ close to 0.5 have relatively high variance. The other major problem is that the predicted probability may lie outside the (0, 1) range and hence produce nonsense probabilities.

Since using linear regression to model a binary outcome can cause violations of the assumptions of the linear regression model, binary probit and logit models are usually used to model binary outcomes with predictor variables that are continuous or categorical. Theses models predict the probability of Y occurring given known values of Xs. By assuming a specific form for the distribution of the error term ε, models can be estimated using maximum likelihood method. Most often, the choice of the error term distribution is between the normal errors which result in the probit model, and the logistic errors which result in the logit model. The function forms of the probit model and logit model is shown in the formula bellow:

Binary Logit Model: $$\Pr(y=1 \mid x) = \frac{\exp(x\beta)}{1+\exp(x\beta)}$$

Binary Probit Model: $$\Pr(y=1 \mid x) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})dt = \phi(x\beta)$$

Model Specification

Table 9.1 provides a list of independent variables used in the model, their definitions, and associated descriptive statistics in the sample. Three categories of the explanatory variables that influence a commuter's route choice propensity, including commute information, primary route attributes, and commuters' socio-demographic information, are included in the empirical analysis.

Table 9.1: Independent Variable Summary

Variable	Definition	Mean	SD
Commute Info			
Trip-chaining stops	Number of stops and drop-offs during the 10 days	3.97	4.55
Dless5	Number of commutes whose departure time deviate less than 5 minutes compared to the median departure time of the 10 days	4.52	2.63
A30more	Number of commutes whose arrival time deviate more than 30 minutes compared to the median arrival time of the 10 days	1.87	2.21
Primary Route Info			
Distance	Average commute distance in miles	15.90	10.84
Travel time	Average travel time (stopping time excluded) in minutes	30.84	15.67
Speed	Average travel speed in mph	29.36	10.13
Duration	Average commute duration (stopping time included) in minutes	37.40	19.63
Number of idle stops	Average number of idles (speed less than 5mph for at least 1 minutes)	2.30	1.47
Percent of freeway	Percentage of freeway travel distance compared to total travel distance	26.93%	29.928%
Number of traffic signals	Number of traffic signals	7.14	4.55
Socio-demographic (dummy variables)			
Gender	Male: reference group	49.45 %	
	Female	50.55 %	
Age group	Between 45 and 52: reference group	26.37%	
	Less than 45	46.15%	
	More than 52	27.47%	
Education	Less than college.: reference group	34.62%	
	College and above	54.95%	
	Unknown	10.44%	
Household income	Income less than $100,000: reference group	54.95%	
	Income larger than $100,000	42.86%	
	Unknown	2.20%	
Residence type	Single house: reference group	89.01%	
	Apartment or townhouse	6.04%	
	Unknown	4.95%	
Tenure at residence	Less than one year: reference group	2.20%	
	One to three years	14.84%	
	More than three years	78.02%	
	Unknown	4.95%	

Individual and Household Socio-demographics

This group of variables is designed to account for the taste variations in choices between different population groups, as well as capturing the effects of the life-cycle stage of a household on route choice behavior. This group of variables includes commuter's age, gender, education level, household size and income, residence type and tenure at residence. Cut-off points for age and income dummy variables were created using tree model analysis based on deviation minimization. The group of commuters with household annual income less than $100,000 is set as the reference group for the income dummy. One dummy variable is used for commuters with household income greater than or equal to $100,000. The group of commuters with age between 45 and 52 is set as the age group dummy. Two dummy variables are used for age group. One dummy variable is for the group of age less than 45, and another is for the group of age larger than 52.

Commute Journey Attributes

This group of variables is designed to capture the impact of work schedule flexibility and trip-chaining on commuters' route choice. Two schedule flexibility variables are developed to reflect the workers' ability to vary their arrival and departure times. The number of commute journeys whose departure time vary less than 5 minutes before or after the median departure time of the ten-day period (Dless5) is chosen to represent departure time inflexibility. The number of commute journeys whose arrival time vary more than 30 minutes before or after the median arrival time of the ten-day period (A30more) represents the arrival time flexibility. The total number of trip-chaining stops made during the study period is used to estimate the trip-chaining frequency.

Primary Route Characteristics

A primary route is the route that a commuter uses most frequently. This group of variables tries to capture the impact of the primary route's traffic condition and driving experience on commuters' decision making. This group of variables includes commute time and distance, average travel speed, number of idle stops, percent of freeways, and number of traffic signal. The reliability of a particular route can be expected to play an important role in the traveler's decision of whether using a secondary route. Travel time standard deviation was proposed to investigate the effect of travel time variability, but due to the small number of observations per route, the travel time standard deviation cannot be taken as a representative value of the travel time variation of a certain commute route. Therefore, it was not included in further model development. Free flow travel time is calculated based on link distance and free flow travel speed of different road functional classes from the Atlanta Regional Commission's transportation planning model. A ratio between the real travel time and the free flow travel time was calculated to represent the congestion level of the primary route, but this variable is not statistically significant in any of the models discussed later. Hence, it is not included in model estimation.

The choice of variables for potential inclusion in the model was guided by previous theoretical and empirical work on route choice modeling. The final specification is based on a systematic process of eliminating variables found not to be statistically significant in previous specifications and based on considerations of parsimony in representation.

Some variables with marginally significant coefficients are retained in the final specification, either for the sake of completeness or because they provide useful and suggestive insights. The univariable models show that among the primary route attributes, average travel speed, percentage of freeway travel distance, and the number of signals have marginal impact on the dependent variable; among the socio-demographic variables, gender, residence type and tenure at residence have marginal impact on the dependent variable. Hence, they are excluded in further model development.

Model Estimation

A correlation matrix was computed to detect potential collinearity between all pairs of the explanatory variables included in model estimation. The resulting correlation coefficients are all less than 0.70, which indicates there is no unacceptable correlation between any two specific variables (Table 9.2).

Table 9.2: Independent variable Correlation Table

	Trip-chaining stops	*Dless5*	*A30more*	*Distance*	*Idle stops*	*Age group1*	*Age group2*	*Income*
Trip-chaining stops	1							
Dless5	-0.2144	1						
A30more	0.1117	-0.5882	1					
Distance	0.3938	-0.1612	0.0308	1				
Idle stops	0.5414	-0.2252	0.0713	0.4183	1			
Age group1	0.0217	-0.1419	-0.0455	0.0367	0.0761	1		
Age group2	-0.0139	0.1974	-0.0635	-0.0471	-0.1039	-0.5501	1	
Income	-0.0085	-0.1219	0.145	0.0542	0.009	-0.1461	0.1014	1

The final model specifications and parameter estimation results are presented in Table 9.3. The first model uses only commute journey information. The second model uses only primary route characteristics. The third model uses only driver and household socio-demographic attributes. The final model uses all three groups of explanatory variables. All the coefficient estimates have the expected signs. Since the coefficient determines the probability that a commuter uses multiple routes, a positive coefficient for a variable means that the probability of using multiple routes increases with the increase in the value of that variable. All the individual coefficient estimates of the first three models except the variable distance are significantly different from zero at the 90 percent confidence level. In the fourth model, all the variables in the first three models are included. The variables including A30more, trip-chaining and dummy for age group younger than 45, are significantly different from zero at the 95 percent confidence level.

Of the three categories of variables discussed in the previous section (commute characteristics, individual attributes, primary route attributes), those describe the characteristics of the commute itself have a dominant effect relative to the other two independent variable categories. This result is consistent with the research result of Mahmassani et al. [1990] in which route switching propensity was based on commuting survey of 638 households.

Table 9.3: Model Estimation Results

Variable	Model 1		Model 2		Model 3		Model 4	
	Coef	*t*	*Coef*	*t*	*Coef*	*t*	*Coef*	*t*
	(Odds)	*(p)*	*(Odds)*	*(p)*	*(Odds)*	*(p)*	*(Odds)*	*(p)*
Commute Info								
Trip-chaining stops	0.2220	4.15					0.2278	3.34
	(1.2486)	(0.000)					(1.2558)	(0.001)
Dless5	-0.1516	-1.77					-0.1490	-1.62
	(0.8593)	(0.077)					(0.8815)	(0.106)
A30more	0.4441	3.15					0.4695	3.08
	(1.5590)	(0.002)					(1.5992)	(0.002)
Primary Route Info								
Distance			0.0177	1.05			0.0086	0.41
			(1.0178)	(0.293)			(1.0087)	(0.680)
Number of idle stops			0.4013	3.05			0.1135	0.68
			(1.4938)	(0.002)			(1.1202)	(0.498)
Socio-demographic								
Age less than 45					-0.7406	-1.86	-1.2730	-2.49
Age larger than 52					(0.4768)	(0.063)	(0.2800)	(0.013)
					-0.7978	-1.86	-0.6550	-1.21
					(0.4503)	(0.072)	(0.5195)	(0.226)
Household income larger than $100,000					0.6747	2.08	0.4696	1.15
					(1.9634)	(0.038)	(1.5994)	(0.252)
Model Summary Statistics								
Constant	-0.2420	0.20	-0.7302	0.028	0.7212	0.038	-0.0332	-0.04
		(0.845)				(2.07)		(0.968)
Log Likelihood at Zero	-121.65		-121.65		-119.29		-118.35	
Log likelihood at convergence	-88.87		-113.07		-114.94		-81.42	
Prob>Chi2	0.0000		0.0000		0.0133		0.0000	
Pseudo R-square	0.2695		0.0705		0.0365		0.3120	
Number of observations	181		181		178		177	

Residual analysis was performed to isolate the points for which the model fits poorly, and isolate points that exert an undue influence on the model. Pearson residuals are the difference between the observation and the fit divided by the square root of the estimated variance for the observation. Leverage or hat value gauges the influences of the observed value of the outcome variable over the predicted values. In average, for normally distributed residuals, 95% should lie between -2 and +2, and 99% should lie between -2.5 and +2.5. Therefore, Pearson residuals with an absolute value greater than 3 may be cause for concern. The average leverage value is defined as (k+1)/n in which k is the number of predictors in the model and n is the number of subjects. Leverage values can lie between 0 and 1. If no cases exert undue influence over the model then all the leverage values should be close to the average value ((k+1)/n). Hoaglin and Welsch (1978) recommended investigating cases with values greater than twice the average (2(k+1)/n) and Stevens (1992) recommended using three times the average (3(k+1)/n) as a cur-off point for identifying cases having undue influence.

If a model fits the sample data well then all residuals should be small. According to the scatter plot of Pearson residual and leverage values shown in Figure 9.1, two sample points have large residuals (Pearson residual has an absolute value larger than 3) and none of the sample points have undue influence on the model (leverage value is larger than the cut off point 0.169). Since no data input error was found in those cases, those two data points remained in the sample.

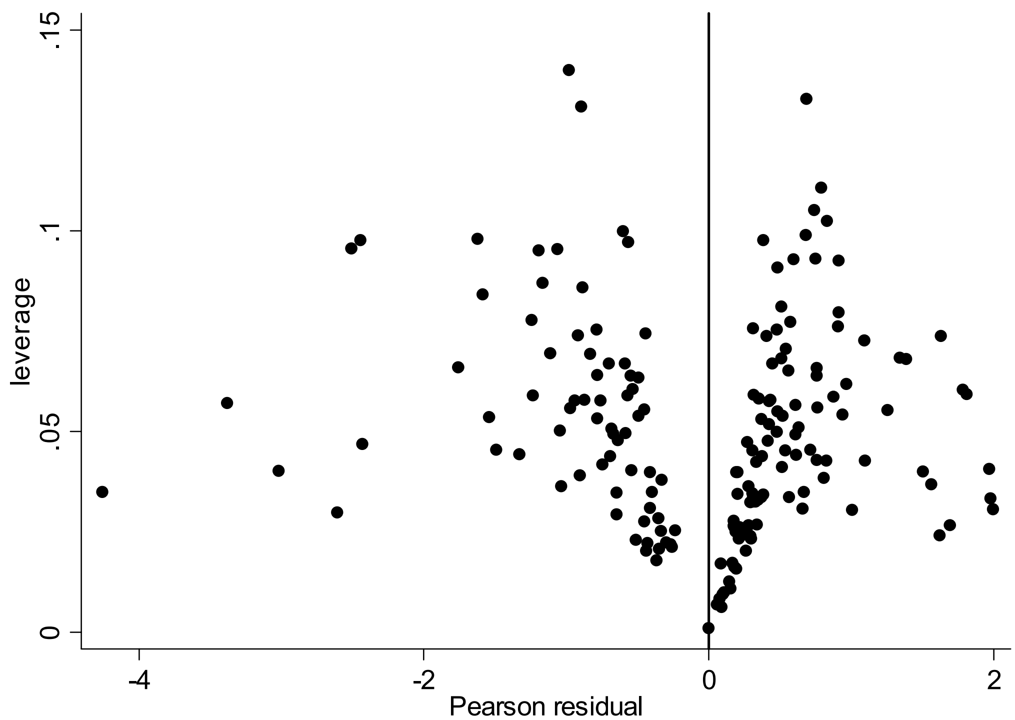


Figure 9.1: Leverage & Pearson Residual of Model 4

Effect of the Explanatory Variables

Changes in the predicted probabilities of using multiple commute routes based on the changes of the independent variables in model 4 are listed in Table 9.4.

Based on model 4, among the commute information variables, trip-chaining and work schedule flexibility increase drivers' propensity of choosing multiple commute routes. The probability of having multiple commute routes is higher for people who make stops during their morning commute than for those who do not make stops. Li et al. [2003] also found similar relationship in the observation of 56 commuters' behavior during a week period. Marginal effect of trip-chaining stops is 0.0474 which indicates an increase of the variable from 0.5 units below to 0.5 units above the mean increases the probability of using multiple routes 4.74%. People with greater schedule flexibility are likely to use multiple commute routes. Based on the model, increasing arrival time flexibility [A30more] can increase probability of using multiple routes. For example, holding all the independent variables at their mean, one unit increase of [A30more] (Number of commutes whose arrival time deviate greater than 30 minutes compared to the median arrival time) will increase the probability of using multiple commute routes by 9.74 percent. The author speculates that this departure time and arrival time flexibility allows commuters to experiment with alternative routes, while not facing the adverse consequences of arriving late for work.

Among the primary route variables, based on model 2, increase of the number of idle stops will increase the probability of choosing multiple commute routes. Based on model 4, both commute distance and the number of idle stops do not have significant impact on commuter's decision of using multiple routes or not. This finding is consistent with Abdel-Aty et al. [1994], who reported that commute distance did not seem to have a significant effect on using alternative routes. Mahmassani et al. [1990] found the propensity to use multiple routes decreased with the increase of average speeds in their study. However, travel speed is not significant in our model based on the univariate model.

Among the socio-demographic variables, based on model 3, the age group dummies (one for age less than 45 and one of age greater than 52) and income dummy (income>$100,000) have significant impact at 90 percent confidence level on the dependent variable. Commuters with higher household income have higher propensity to choose multiple commute routes. Age group 45 to 52 have higher propensity to choose multiple routes compare to the age group younger than 45 and age group older than 52. In model 4, the dummy variable of the age group less than 45 remained significant.

Abdel-Aty et al. [1994] also found a correlation between income and using alternative routes in their study; the fraction of individuals with alternative routes (percent of multiple route users within each income category) increases from 6.7 percent among those with incomes less than $25,000 to 28 percent among those with incomes more than $100,000 in their sample. Abdel-Aty et al. [1994] found the same relationship for level of education: highly educated people tended to use alternative routes.

Gender is not significant in our model. This is inconsistent with previous research. Mannering and Kim [1994] and Mannering [1989] reported that men were more likely to change routes than women. Familiarity of network is expected to have influence on commuters' propensity of using alternative routes. Assuming tenure of residence as an indicator of familiarity of the area, researchers expect that longer tenure of residence will indicate higher propensity of using alternative routes, but since most drivers in the sample have been living in the current location for more than three years, the effect of this variable was not evident in this sample.

Table 9.4: Changes in Predicted Probabilities

	from:	to:	dif:	from:	to:	dif:	from:	to:	dif:	from:	to:	dif:	
	x=min	x=max	min-> max	x=0	x=1	0->1	x-1/2	x+1/2	x-1/2-> x+1/2	x-1/2sd	x+1/2sd	x-1/2sd-> x+1/2sd	Marginal Effect
Trip-chaining stops	0.5071	0.9899	0.4828	0.5071	0.5637	0.0566	0.6808	0.7281	0.0473	0.5898	0.7989	0.2091	0.0474
Dless5	0.8233	0.5121	-0.3112	0.8233	0.8006	-0.0227	0.7203	0.6893	-0.031	0.744	0.6628	-0.0812	-0.031
A30more	0.497	0.9908	0.4939	0.497	0.6124	0.1154	0.6539	0.7514	0.0974	0.5873	0.8005	0.2133	0.0976
Distance	0.677	0.7661	0.0891	0.676	0.6779	0.0019	0.7041	0.7059	0.0018	0.6954	0.7144	0.019	0.0018
Number of idle stops	0.6483	0.794	0.1457	0.6483	0.6737	0.0254	0.6931	0.7167	0.0236	0.6873	0.7221	0.0347	0.0236
Age less than 25	0.8095	0.5433	-0.2662	0.8095	0.5433	-0.2662	0.8187	0.5584	-0.2603	0.7665	0.6349	-0.1316	-0.2648
Age larger than 52	0.7413	0.5981	-0.1432	0.7413	0.5981	-0.1432	0.7683	0.6327	-0.1356	0.7346	0.6735	-0.0611	-0.1362
Household income higher than $100,000	0.6602	0.7566	0.0963	0.6602	0.7566	0.0963	0.6539	0.7514	0.0975	0.6801	0.7287	0.0486	0.0977

Chapter 10

Conclusions

This chapter discusses the implications of this dissertation. The first section discusses the contributions of this research effort. The second section offers some suggestions for future research needs.

Summary of Research Findings

The research efforts reported in this dissertation includes studies of the spatial route deviation patterns and the factors that influence morning commuters' route choice and route switching based on objective real-world observations of travel behavior during multi-day period.

This dissertation studies the spatial pattern of route choice that was impossible to discern with earlier conventional survey methods. The dissertation defined a group of eight deviation patterns. This research found that commute distance plays an important role in commuters' route deviation pattern. When commute distance are relatively long, most commuters only deviate near the home end possibly due to the less familiarity of the remaining part of the network. This dissertation also found that deviations along the middle of the routes share lower distance percentage comparing to deviations close to trip origins and destinations.

Investigation of the objective route choice factors confirmed that minimizing travel time, although very important, is not the only factor that impacts route choice. Several other factors have been identified that appear to significantly impact commuters' route choice. These factors include driving experience such as the number of idle stops and traffic signals on the way, road functional classification such as percentage of freeways. Drivers tend to choose routes with shorter travel time and distance, consisting larger freeway percentages, less traffic signals, and less idle stops as their primary commute routes.

Models of the objective route attributes can explain around 14 percent of the variability in the sample set. The unexplained part may be attributed to factors such as differences in perception, imperfect information on route costs. To some degree, the drivers' knowledge of attribute values is a distorted image of the actual values. The travelers' perception of relevant alternatives and their attributes is somewhat incomplete and inaccurate. Subjective perceptions and attitudes about objective route attributes on different alternatives drive them to a certain choice. Different usage of route attributes, different perceptions of route attributes, different interpretation of the traffic network situation can result in different behavior even in the same situation.

This dissertation examined the choice of using single or multiple morning commute routes. The results indicate the strong explanatory power of work schedule flexibility and trip-chaining on the dependent variable comparing to the commuters' socio-demographic characteristics and commute route related attributes. Among the commute information variables, trip-chaining and work schedule flexibility increase drivers' propensity of choosing multiple commute routes. The probability of having multiple commute routes is higher for people who stop during their morning commute than for those who do not stop. People with greater schedule flexibility are likely to use multiple commute routes. Among the socio-demographic variables, age and income have significant impact on the dependent variable. Commuters from the age group between 45 and 52 have a higher tendency to use multiple routes than people from other age groups. Higher income participants tended to use alternative routes (based on model 3 in chapter 9). Gender was not significant in our model. Among the primary route variables, commute distance does not have significant impact on commuter's decision of using multiple routes or not.

A better understanding of route choice is important in improving traffic assignment methods that are one of the major transportation planning modeling steps. These findings are useful in generating more realistic general route cost functions. A function that takes a combination of all the factors identified previously after assign each with the appropriate weight according to its significance would be a more realistic approach and can allocate trips to the appropriate road segments in the traffic assignment. Better understanding of the route choice behavior can also help design Advanced Traveler Information Systems (ATIS) algorithms that generate routes based on assumptions of travel time minimization with other considerations of driving experience such as traveling on freeways or local streets, and the number of traffic control devices on the way. These algorithms will have more appeal to ATIS customers.

Recent developments of planning models are based on simulation of the daily activities of individual travelers. These models attempt to capture the activities, decisions, and spatial motion of travelers through time at the individual level. The discrete-level research findings in this dissertation can be used to improve the route choice model assumptions used in the agent-based transportation simulation models such as, one of the most advanced modeling approaches, TRANSIMS.

The focus of today's transportation planning is increasingly moving away from the transportation investments that meet unrestricted demand, to the applications of new technologies that manage travel demand and achieve more efficient use of the systems. Understandings of travelers' route choice behavior are central to the modeling of travel behavior and the assessment of policy impacts.

Most research on route choice behavior is based on survey methods either revealed preference survey or stated preference survey, or laboratory simulations that repeatedly ask the participants to respond to hypothetical route choices. In contrast to previous research, the work reported here is based on real data of drivers' choices from field observations. This dissertation presents an extensive effort in analyzing GPS-based travel behavior data. One of the contributions of this dissertation is the development of methodologies to extract route choice information and trip-level travel information from the GPS-based vehicle activity data. These findings are important to incorporating GPS-based data into the traditional travel survey.

Recommendations on Future Work

Although GPS data provide a very accurate record of travel behavior, GPS data by themselves do not provide information about the underlying reasons why travelers choose a certain route over the others. Travelers' decision-making process, their perceptions and knowledge about these routes are unknown in this study. Investigation into how much information drivers have about their routes, their awareness of alternate routes, their usage of different traffic information either before or during the trip, traveler's normal travel patterns such as day-to-day behavior (work schedule, route choice and response to recurring congestion), pre-trip and en-route response to unexpected congestion information, delay tolerance threshold, willingness to change driving patterns would be helpful to further studies on route choice behavior.

A study that combines both the field observation of travel behavior and survey methods that record the traveler's decision making process can provide more insightful discovery on travelers' route choice behavior. An example of this comprehensive approach can be found in the ongoing study of Doherty et al. [2004], which combines GPS and GIS technologies with a recently developed computerized activity scheduling survey that has the potential to simultaneously observe detailed spatial-temporal activity-travel patterns and underlying decision processes of individuals within a household over long periods of time, while at the same time minimizing respondent burden.

The reliability in aspect of travel time of a particular route can be expected to play an important role in the traveler's route choice behavior [Jackson and Jucker, 1981], [Abdel-

Aty, 1995]. Researchers intended to use travel time standard deviation to investigate the effect of travel time variability. This study did not estimate travel reliability's impact on morning commuters' route choice at discrete level due to data limitation because the small number of observation per route, the travel time standard deviation cannot be taken as a representative value of the travel time variation of a certain commute route. A possible extension for the work is to expand the study period and get reliable estimations of travel time reliability information for each route.

Based on the size and study duration of this instrumented vehicle activities available in the Commute Atlanta research, future research can accurately estimate the typical travel speed and link travel time on freeways and major roads during a certain time-of-day. A research that compares the shortest-time route identified based on link travel time and the actual routes utilized by the commuters can provide more insight on the route choice behavior.

REFERENCES

Abde-Aty, M. A., K. M. Vaughn, R. Kitamura, P. P. Jovanis, and F. L. Mannering (1994). Models of Commuters' Information Use and Route Choice: Initial Results Based on Southern California Commuter Route Choice Survey. Transportation Research Record 1453

Abde-Aty, M. (1995). Investigating the Factors Influencing Route Choice: New Approaches. Disseration. University of California, Davis

Abdel-Aty, M. A., R. Kitamura and P. P. Jovanis (1995). Investigating Effect of Travel Time Variability on Route Choice Using Repeated-Measurement Stated Preference Data. Transportation Research Record 1493, pp.39-45

Abdel-Aty, M. A., R. Kitamura and P. P. Jovanis (1996). Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice. Transportation Research Part C PII: S0968-090X (96)00023-X

Abu-Eisheh, S. A. and F. L. Mannering (1987). Discrete/Continuous Analysis of Commuters' Route and Departure Time Choice. Transportation Research Record 1138

Antonisse, R., A. Daly, and M. Ben-Akiva (1989). A Highway Assignment Method Based on Behavioral Models of Car Drivers' Route Choice. Transportation Research Record 1220, TRB, National Research Council, Washington, D.C.

Axhausen, K.W., S. Schonfelder, J. Wolf, M. Oliveira and U. Samaga (2003). Eighty Weeks of GPS Traces: Approaches to Enriching Trip Information. 83rd Transportation Research Board Meeting CD-ROM

Ben-Akiva, M., A.D. Palma and I. Kaysi (1991). Dynamic network models and driver information systems. Transportation Research Part A 25A (5), pp.251-266

Bianco, M. and C. Lawson (1996). Trip Chaining, Childcare, and Personal Safety. Women's Travel Issues. Proceedings from the Second National Conference, October 1996, pp.124

Bovy, P. and E. Stern (1990). Route Choice: Wayfinding in Transport Network.

Bunch, D., M. Bradley, T. Golob, R. Kitamura and G. Ochiuzzo (1993). Demand for Clean-Fuel Vehicles in California: A Discrete-Choice Stated Preference Pilot Project, Transportation Research A, Vol. 27A (3), pp.237-253

Cascetta, E., A. Papola, F. Russo and A. Vitetta (1999). Implicit Availability / Perception Logit Models for Route Choice in Transportation Networks. 8th WCTR proceedings, volume 3

Cascetta, E., F. Russo, F. A. Viola and A. Vitetta (2002). A Model of Route Perception in Urban Road Networks. Transportation Research Part B 36 pp.577-592

Davidson, D. (1991). Impact of Suburban Employee Trip Chaining on transportation Demand Management. Transportation Research Record 1321, TRB, National Research Council, Washington, D.C., pp. 82-89

Dia, H. (2002). An Agent-Based Approach to Modeling Driver Route Choice, Behavior Under the Influence of Real-Time Information. Transportation Research Part C 10, pp.331-349

Doherty, S. T., N. Noel, M. Gosselin, C. Sirois, M. Ueno and F. Theberge (2004). Moving Beyond Observed Outcomes: Integrating Global Positioning Systems Computer-Based Travel Behaviour Surveys. Transportation Research Circular: Personal Travel: The Long and Short of It. Transportation Research Board, National Research Council, Washington, D.C. (in preparation)

Domenchich, T. and D. McFadden (1975). Urban Travel Demand: A Behavioural Analysis. North Holland, Amsterdam.

Greenfeld, J. (2002). Matching GPS Observations to Locations on a Digital Map, proceedings of 81st TRB conference

Hanson, S. (1980). The Importance of the Multi-purpose Journey to Work in Urban Travel Behavior. Transportation 9, pp.229-248

Hoaglin, D. and R. Welsch (1978). The hat matrix in regression and ANOVA. American Statistician, 32, pp.17-22

Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins (1992). GPS: Theory and Practice. Third Edition, Springer-Verlag Wien New York

Hosmer, D. W. and S. Lemeshow (2000). Applied Logistic Regression. Second Edition, ISBN0-471-35632-8

Hsiao, C. (1996). Chapter 16 Logit and Probit Models. The Econometrics of Panel Data: A Handbook of the Theory with Applications, Second Edition

Jackson, W. B. and J. V. Jucker (1981). An Empirical Study of Travel Time Variability and Travel Choice Behavior. Transportation Science, Vol. 16, No. 4, November 1981

Jan, O., A. Horowitz and Z. Peng (2000). Using Global Positioning System Data to Understand Variations in Path Choice. Transportation Research Record 1725, TRB, National Research Council, Washington, D.C., pp.37-44.

Journey to Work Trends: in the United Statesand its Major Metropolitan Areas (1960 – 2000). (2004). Publication No. FHWA -EP-03-058. http://www.fhwa.dot.gov/ctpp/jtw/index.htm

Khattak, A. J., J. L. Schofer, and F. S. Kopplman (1991). Factors Influencing Commuters' En Route Diversion Behavior in Response to Delay. Transportation Research Record 1318

Khattak, A. J., J. L. Schofer and F. S. Koppelman (1995). Effect of Traffic Information on Commuters' Propensity to Change Route and Departure Time.

Kim, S. and J.-H. Kim (2001). Adaptive Fuzzy-Network-Based C-Measure Map-Matching Algorithm for Car Navigation System, IEEE Transactions on Industrial Electronics, Vol. 48, No. 2, April 2001, pp.432-441.

Lam, T. (2000). Route and Scheduling choice Under Travel Time Uncertainty. Transportation Research Record 1725, p71-78, Paper No. 00-0923

Lam, T. C. and K. A. Small (2001). The Value of Time and Reliability: Measurement from a Value Pricing Experiment. Transportation Research Part E 37, pp.231-251

Lamb P. and S. Thiebaux (1999). Avoiding Explicit Map-Matching In Vehicle Location, http://discus.anu.edu.au/~thiebaux/

Li, H., R. Guensler, J. Ogle and J. Wang (2004). Using GPS data to understand day-to-day dynamics of morning commute behavior. 83th TRB Annual Meeting CD

Lockwood, P. B. and M. J. Demetsky (1994). Nonwork travel – A study of changing behavior, presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, DC, January.

Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables, ISBN: 0803973748

Long, J. S. and J. Freese (2001). Regression Models for Categorical Dependent Variables Using Stata. ISBN 1-881228-62-2

Lotan, T. (1996). Effects of Familiarity on Route Choice Behavior in the Presence of Information. Transportation Research Part C PII: S0968-090X(96)00028-9

Louviere J. and G. Woodworth (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregated Data, Journal of Marketing Research, pp.350-367

Mahmassani, H. S. and D. G. Stephan (1988). Experimental Investigation of Route and Departure Time Choice Dynamics of Urban Commuters. Transportation Research Record 1203

Mahmassani, H. S., C. G. Caplice and C. M. Walton (1990). Characteristics of Urban Commuter Behavior: Switching Propensity and Use of Information. Transportation Research Record 1285

Mahmassani H., S. Hatcher, and C. Caplice (1996). Daily Variation of Trip Chaining, Scheduling and Path Selection Behavior of Work Commuters. Understanding Travel Behaviour in an Era of Change.

Makimura K., H. Kikuchi, S. Tada, Y Nakajima, H. Ishida andd T. Hyodo (2002). Performance Indicator Measurement Using Car Navigation Systems, proceedings of 81st TRB conference

Mannering F. (1989). Poisson Analysis of Commuter Flexibility in Changing Route and Departure Time. Transportation Research 23B(1), pp.53-60.

Mannering, F. and S. G. Kim (1994). Statistical Analysis of Commuters' Route Choice, Mode, and Departure Time Flexibility. Transportation Research C. Vol. 2. No. 1. pp.35-47

McFadden, D.J. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, in Frontiers in Econometrics ed. By P. Zarembka, New York: Academic Press

McFadden, D. (1981) "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden (eds), STRUCTURAL ANALYSIS OF DISCRETE DATA, MIT Press: Cambridge, 198-272.

Meyer, D. M. and E. J. Miller (1984). Urban Transportation Planning: A Decision-Oriented Approach.

Nakayama, S. and R. Kitamura (2000). Route Choice Model with Inductive Learning. Transportation Research Record 1725, p63-70, Paper No. 00-0727

Nishii K., K. Kondo, and R. Kitamura (1989). Empirical Analysis of Trip Changing Behavior. Transportation Research Record 1203, TRB, National Research Council, Washington, D.C., pp.49

Ogle, J., Guensler, R. and Elango V. (submitted 2004). Commute Atlanta Value Pricing Program: Recruitment Methods and Travel Diary Response Rates; Transportation Research Record; National Academy of Sciences; Washington, DC.

Orski K. (1989). A Realistic Appraisal of Traffic Congestion. Urban Land, Oct. 1989, pp. 35

Ortuzar, J. D. D. and L. G. Willumsen (1990). Modelling Transport. published by John Wiley & Sons Ltd., ISBN 0471926299

Parkinson W. B., and James J. (1996). Global Positioning System: Theory and Applications. Amer Inst of Aeronautics, ISBN: 1563471078

Polydoropoulou, A., M. Ben-Akiva, And I. Kaysi (1994). Influence of Traffic Information on Drivers' Route Choice Behavior. Transportation Research Record 1453, p56-65

Purvis, C. L. (1994). Changes in regional travel characteristics and travel time budgets in the San Francisco bay area: 1960 – 1990. Presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, DC, January.

Rabe-Hesketh, S. and B. Everitt (2004). A Handbook of Statistical Analyses using Stata, Third Edition. ISBN 1-58488-404-5

Richard, H., M. Emmerink, P, Nijkamp, P. Rietveld, Jos N. andV. Ommeren (1995). Variable message Signs and Radio Traffic Information: An Integrated Empirical Analysis of Drivers' Route Choice Behavior. Transportation Research Part A 0965-8564(95)00016-X

Scott C. (1994). Improved GPS Positioning for Motor Vehicles Through Map Matching, Vehicle Navigation Information Systems Conference (VNIS 94), pp.1391-1400.

Shbaklo, S. A., F. S. Koppelman and J. L. Schofer (1995). Measures of Route Similarity and Compliance: Application to ATIS. Transportation Congress

Sheffi Y. (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. 1985 by Prentice-Hall, Inc. ISBN: 0-13-939729-9-01

Srinivasan, K. K. and H. S. Mahmassani (2000). Modeling Inertia and Compliance Mechnisms in Route Choice Behavior Under Real-Time Informaiton. Transportation Research Record 1725, p45-53, Paper No. 00-0568

Stern, E., J. Tzelgov and A. Henik (1983). Driving Efforts and Urban Route Choice. The Logistics and Transportation Review, volume 19

Stern, E., E. Holm and M. V. Maarseveen (1993). Information and Commuters Behaviour: A Comparative Analysis. Europe on the Move, pp.141-155

Stevens, J. P. (1992). Applied multivariate statistics for the social sciences (2nd edition). Hillsdale, NJ: Erlbaum.

Taylor G. and G. Blewitt (1999). Virtual Differential GPS & Road Reduction Filtering by Map Matching, Abstract for ION99

Wardrop, J.G. (1952). Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineering. Part II 1, pp.325-262.

Washington, S. P., M. G. Karlaftis and F. L. Mannering (2003). Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC, 2003

White C. E., D. Bernstein and L. Kornhauser (2000). Some Map Matching Algorithms For Personal Navigation Assistants, Transportation Research Part C 8, pp. 91-108.

Wolf J., S. Hallmark, M. Oliveira, R. Guensler and W. Sarasua (1998). Accuracy Issues with Route Choice Data Collection by Using Global Positioning System. In Transportation Research Record 1660, TRB, National Research Council, Washington, D.C., 1998, pp. 66-74.

Wolf, J. (2000). Using GPS Data Loggers To Replace Travel Diaries In the Collection of Travel Data. Ph. D. thesis, Georgia Institute of Technology

Yalamanchili L., R. Pendyala, N. Prabaharan, and P. Chakravarthy (1999). Analysis of Global Positioning System-Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior. Transportation Research Record 1660, TRB, National Research Council, Washington, D.C., 1999, pp.58-65

Yamamoto, T., S. Fujii, R. Kitamura and H. Yoshida (2000). Analysis of Time Allocation, Departure Time, and Route Choice Behavior Under Congestion Pricing. Transportation Research Record 1725, p95-101, Paper No. 00-0710

Yim Y.B., Youngbin, and R.Cayford (2002). Positional Accuracy of Global Positioning System and Cellular Phone Tracking for Probe Vehicles, proceedings of 81st TRB conference

Yim Y.B., Youngbin, and R. Cayford (2000). Investigation of Vehicles as Probes Using Global Positioning System and Cellular Phone Tracking: Field Operational Test, California PATH Working Paper, UCB-ITS-PWP-2001-9

Appendix A

Accuracy check of the traffic signal information in the RC database

The author compared the traffic signal information in the RC database in the City of Atlanta with the signal control map from the traffic & transportation office of the City of Atlanta. Among the 878 signals in the City of Atlanta, 716 (87.53%) of them were accurately represented in the RC database. The remaining 162 (12.47%) were missing in the RC database.

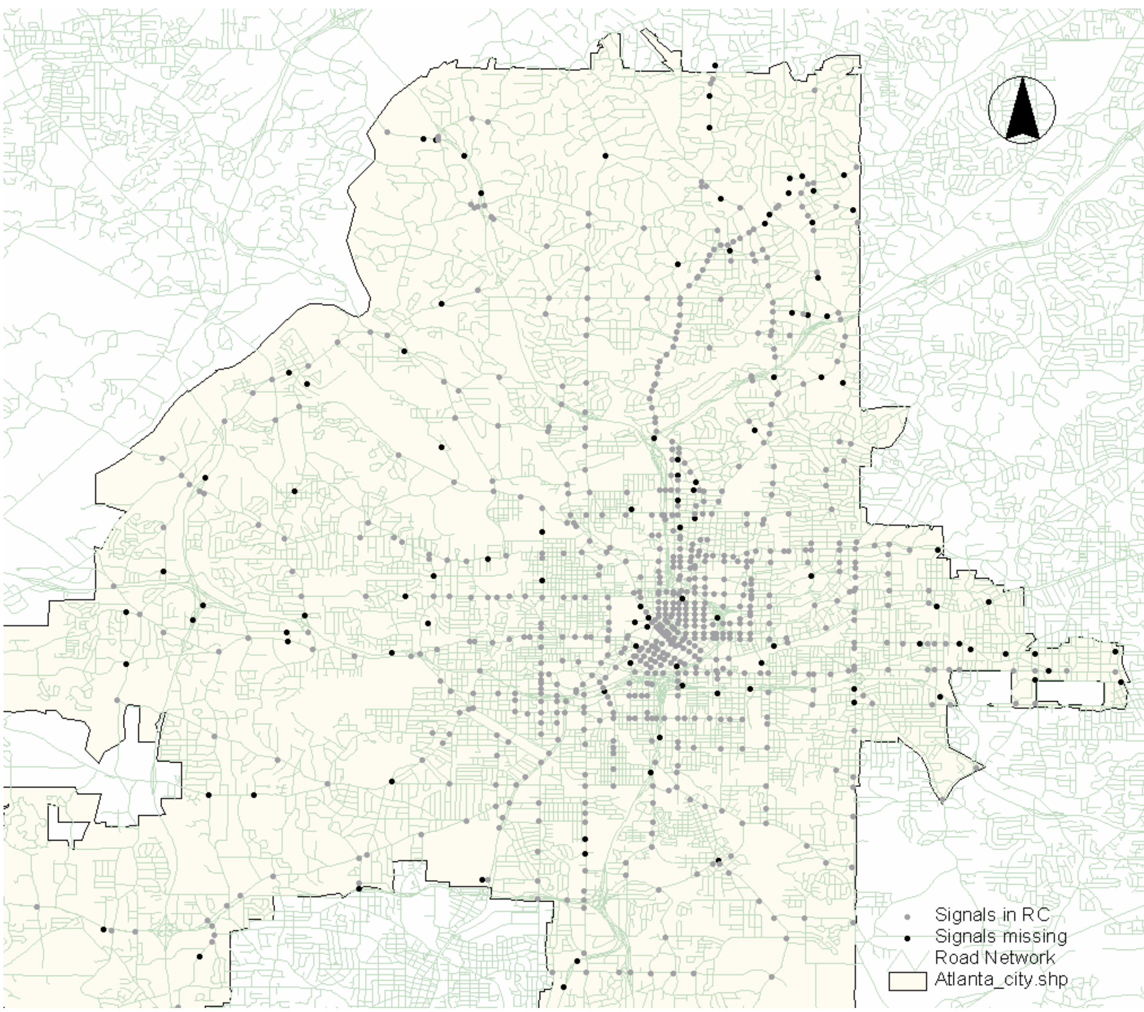


Figure A.1: Signals in the City of Atlanta (signals in black are missing in the RC database)

The author also did random check of traffic signals in three sites of Dekalb, Cobb and Gwinnett counties. Among the 20 signals in the three locations, 18 (90%) of them were accurately represented in the RC database, and the remaining 2 (10%) were missing. The results are listed below (signals in circles are missing in the RC database):

Figure A.2: Site in Cobb County

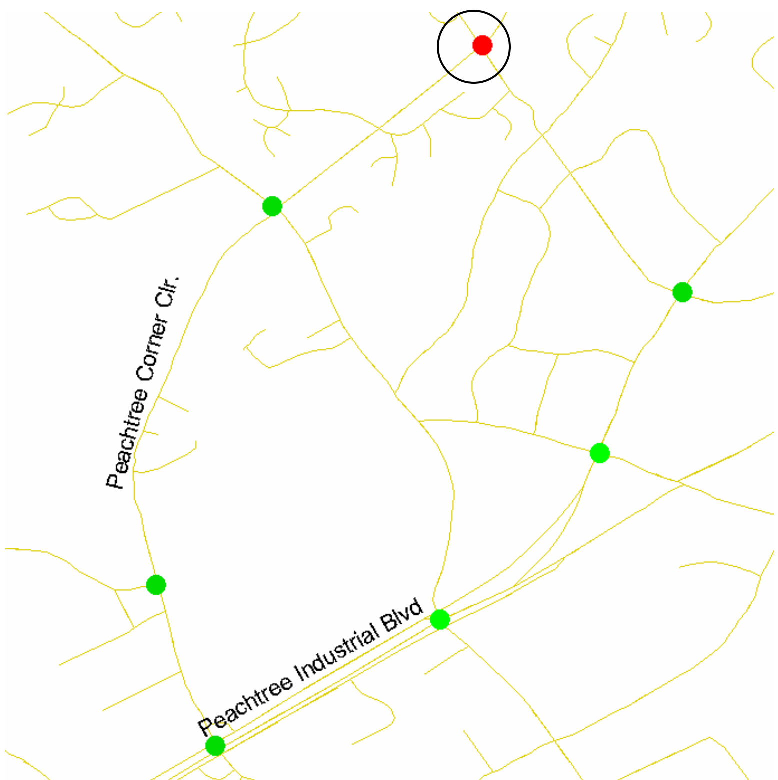


Figure A.3: Site in Gwinnett County

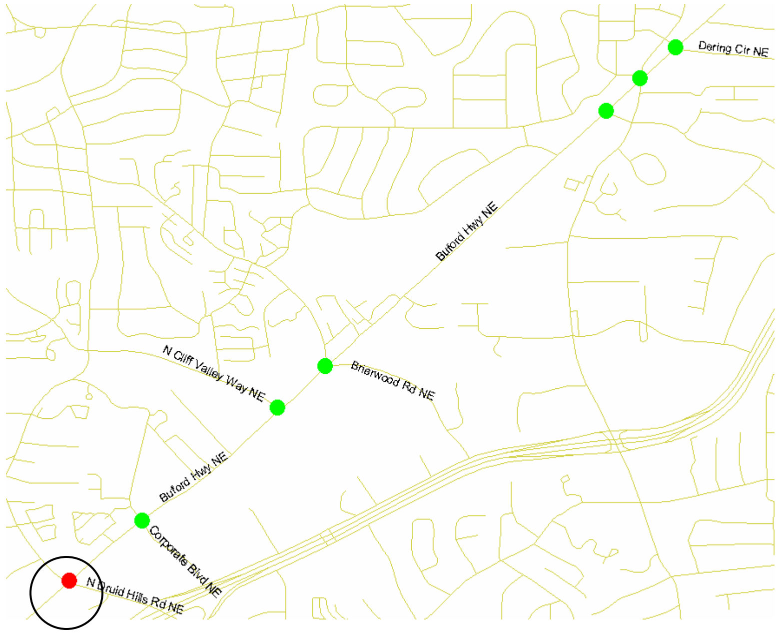


Figure A.4: Site in Dekalb County